Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry with Chemometric Analysis

Amanda E. M. Sinha

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2004

Program Authorized to Offer Degree:
Department of Chemistry

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

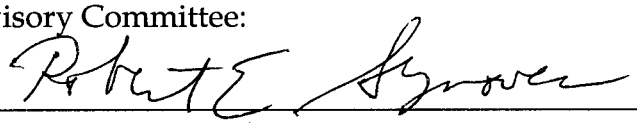
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Amanda E. M. Sinha

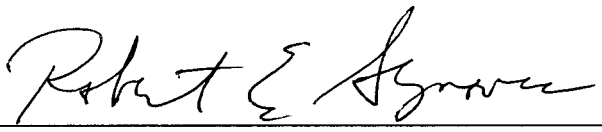
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

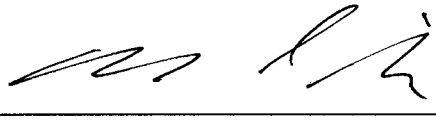


Robert E. Synovec

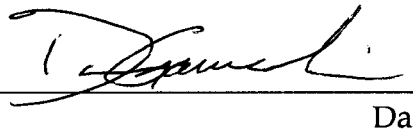
Reading Committee:



Robert E. Synovec



Norman J. Dovichi



Daniel R. Gamelin

Date: August 6, 2004

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Proquest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Amanda E.M. Simba

Date August 6, 2004

University of Washington

Abstract

Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry with Chemometric Analysis

Amanda E. M. Sinha

Chair of the Supervisory Committee: Professor Robert E. Synovec
Department of Chemistry

Multidimensional separations, in particular Comprehensive Two-Dimensional Gas Chromatography (GC x GC), have proven to be powerful techniques that have found a niche in complex mixture analysis. Instruments of this type produce second-order data that is applicable to chemometric analysis. This combination of multidimensional separations and chemometrics results in reduced analysis times by relaxing the chromatographic resolution requirements for quantitation. By coupling GC x GC to a Time-of-Flight Mass Spectrometer (TOFMS) a truly third-order technique is produced. One separation on a GC x GC-TOFMS provides retention times on two chromatographic columns and a complete mass spectrum for each component of a mixture. Using the added selectivity of the mass spectrometer combined with chemometric techniques, overlapping components in a complex mixture can theoretically be deconvoluted using only one data set. Second-order data combined with chemometric techniques require a sample and a standard data set to deconvolute overlapping analytes.

In this dissertation, aspects of GC x GC and GC x GC-TOFMS combined with chemometrics are presented. A novel configuration of a valve-based GC x GC is proposed and tested to extend the workable temperature

range of the valve-based instrument to analyze semi-volatiles. The data structure was confirmed to be applicable to chemometric analysis. This new configuration is combined with a TOFMS to produce the first valve-based GC x GC-TOFMS that is subsequently evaluated for performance. A methodology involving PARAFAC is developed to deconvolute partially overlapped peaks in this data. The chemometric analysis is shown to work for both valve-based and cryogenically modulated GC x GC-TOFMS instruments. The logistics and performance characteristics of peak deconvolution with PARAFAC are analyzed with a set of butyl benzene isomers. The reproducibility and accuracy of deconvolution is subsequently tested as a function of multivariate selectivity, which is related to chromatographic resolution and mass spectral similarity. The deconvolution results for PARAFAC are compared with those for two common GC-MS deconvolution programs. In addition, a tool to locate compounds or classes of compounds of interest based on mass spectral similarity to aid in the analysis of complex mixtures is presented.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number **Page**

LIST OF ABBREVIATIONS

%RSD Percent Relative Standard Deviation

1D One-Dimensional

1D-R_s One-Dimensional Chromatographic Resolution

2D Two-Dimensional

2D-PAGE Two-Dimensional Polyacrylamide Gel Electrophoresis

2D-Rs Two-Dimensional Chromatographic Resolution

3D Three-Dimensional

AMDIS Automated Mass Spectral Deconvolution and Identification System

ANOVA Analysis of Variance

BB n-Butyl Benzene

Bkgd Background

BSTFA N,O-bis[Trimethylsilyl]trifluoroacetamide

CE Capillary Electrophoresis

CE x CE Comprehensive Two-Dimensional Capillary Electrophoresis

ClBz Chlorobenzene

Col. Column

DMMP Dimethyl Methane Phosphonate

DMP Dimethyl Phosphite

Eq. Equation

FID Flame Ionization Detector

FOM Figures of Merit

GC Gas Chromatography

GC x GC Comprehensive Two-Dimensional Gas Chromatography

GC x GC-TOFMS Comprehensive Two-Dimensional Gas Chromatography – Time-of-Flight Mass Spectrometry

GC/MS Gas Chromatography/Mass spectrometry

GRAM Generalized Rank Annihilation Method

H High-Resolution Case

I Intermediate-Resolution Case

i.d. Inner Diameter

IB Iso-Butyl Benzene

Inj. Conc. Injected Concentration

L Low-Resolution Case

LC Liquid Chromatography

LC/MS Liquid Chromatography/Mass Spectrometry

LC x CE Comprehensive Liquid Chromatography - Capillary Electrophoresis

LC x GC Comprehensive Liquid Chromatography - Gas Chromatography

LOD Limit of Detection

m/z Mass-to-Charge Ratio

MPA 1-Methoxy-2-Propyl Acetate

MS Mass Spectrometer

MSD Mass Selective Detector

nas Net Analyte Signal

o.d. Outer Diameter

PARAFAC Parallel Factor Analysis

PCA Principal Component Analysis

PEG Polyethylene Glycol

$\mathbf{R_s}$ Chromatographic Resolution

$\mathbf{R_t}$ Retention Time

SB Sec-Butyl Benzene

SEL Multivariate Selectivity

S/N Signal-to-Noise

TB Tert-Butyl Benzene

TEPT Triethyl Phosphorothioate

TIC Total Ion Current

TLC Thin Layer Chromatography

TLC x TLC Comprehensive Two-Dimensional Thin Layer Chromatography

TLD Trilinear Decomposition

TMS Trimethylsilyl

TOFMS Time-of-Flight Mass Spectrometry

Tri-PLS Trilinear Partial Least Squares

ACKNOWLEDGEMENTS

There are many people who have made this journey a bit easier and more enjoyable. I am unsure how I can thank everyone enough, but I will try. I would especially like to thank my research advisor, Professor Robert E. Synovec, for providing guidance and support over the course of my graduate education. In addition, I would like to thank Dr. Kevin J. Johnson and Dr. Bryan J. Prazen for all that they have done for me, from teaching me my first Matlab commands, to providing insightful discussions, to editing papers and much, much more- Thank you so much! I would also like to thank the other current and former Synovec Group members, especially Bethany Staggemeier, Gwen Gross, Dr. Colin Costin, Janiece Hope, and Adam McBrady, who are some of the most gifted people with whom I could ever hope to work.

I am grateful for having the opportunity to meet and befriend some amazing people in graduate school who are not merely exceptional chemists, but truly great people as well. A special thanks to Jennifer Seymour, Bethany Staggemeier, Gwen Gross, and Gina Huset for lunches, good times, listening, and helping keep me sane over the years - not sure I would have made it without you!

I am deeply indebted to my parents, Barbara and Royal Moses, for all that they have done for me over the years. I am not sure that more loving, generous, supportive people than they exist in this world. I am amazingly lucky to have you as parents - thanks so much for everything! I would also like to thank my brother, Christian Moses, and Gemma Alexander for making my transition to Seattle easier with their presence and providing me with surrogate pets. By the way, I am still sorry about that tennis ball.

I cannot name everyone here, but I would also like to thank my extended family, especially my grandmother, Bertha Soley, and my aunts and

uncles, Betty and Steve Johnson and Dottie and Dewey Reed, for all their thoughtfulness, generosity, and support over the years. I would like to express my sincere gratitude for my mother- and father-in-law, Mita and Provat Sinha, for being the kind and accepting people that they are. I am also indebted for them for enabling my "mini-sabbatical" trip to India during my studies. It was an experience I shall never forget - I am humbly grateful. I would also like to thank Paramita Babli Sinha, for helping to put graduate school in perspective and for being a friend as well as my sister-in-law. Most importantly, I would like to thank my husband, Parikhit Ricky Sinha, for everything he does and has taught me. I cannot fully express how grateful I am for his calming influence, unwavering support, and ability to put things in the right perspective. I would like to thank him for teaching me the mantra that has kept me going and ultimately gotten me to this point: You just have to get *something* done everyday.

DEDICATION

To the memory of my grandfather,

Dr. Russell Clyne Soley.

CHAPTER 1. Chemometric Analysis of Multidimensional Separations*

1.1 Introduction

The desire to understand the chemical composition of complex mixtures is at the heart of many applications in separation science. According to Giddings, "a separation process must entail a selective movement through space of one component with respect to another" [Giddings 1991]. Chromatography is a form of separation in which a stationary phase selectively retains some compounds more than others while a mobile phase transports the less retained analytes away from those that are retained. In capillary gas chromatography, the mobile phase is a carrier gas such as helium or hydrogen and the stationary phase is typically a polymer coating on the capillary wall. High-resolution chromatography using open-bore fused silica capillary columns has dramatically improved gas chromatography (GC) as a means to separate components of complex mixtures [Kaiser 1985]. This technique has wide applicability for analyzing a myriad of complex mixtures in drug development and pharmaceuticals, environmental studies, the petroleum industry, clinical chemistry, pesticides and residues, and foods [Grob 1985]. Some mixtures entail such levels of complexity that the separation of all components, if possible at all, only occurs after very long chromatographic run times.

One option for reducing chromatographic run times for separations of complex mixtures is to combine two or more separation techniques comprehensively. Since Giddings laid out the basic concepts and theory

---

* Portions of this Chapter have been previously published [Sinha et al. 2003a; Sinha et al. 2004d].

behind comprehensive 2D separations nearly two decades ago [Giddings 1984; Giddings 1987], there have been continuous developments in this field. Essentially, during a 2D chemical separation a sample of interest is subjected to separation via two complementary mechanisms. Two-dimensional separations have a variety of manifestations including TLC x TLC [Guiochon et al. 1982], LC x LC [Holland and Jorgenson 2000], LC x CE [Moore and Jorgenson 1995], CE x CE [Michels et al. 2002], LC x GC [Quigley et al. 2000], GC x GC [Dallüge et al. 2003; Ledford et al. 1996], 2D-PAGE [Harris et al. 2002] and MudPIT analyses [Washburn et al. 2002; Washburn et al. 2003]. The term "comprehensive" is used with 2D separations when every component of the sample is representatively subjected to both separation mechanisms such that the resolution gained by the first dimension is maintained during the second dimension separation [Shoenmakers et al. 2003]. Separating in two dimensions simultaneously can shorten run times considerably compared to 1D separations and affords selectivity not available with a 1D separation. The peak capacity in 2D separations can be as high as the product of the peak capacity in each dimension. Hence, these techniques have found a niche for complex mixture analysis.

One of the most common 2D separation techniques is comprehensive two-dimensional gas chromatography (GC x GC). This technique has been demonstrated extensively to be well-suited to analyzing complex chemical mixtures [Beens et al. 2001b; Bertsch 2000; de Geus et al. 1998; de Geus et al. 2001; Dimandja et al. 2000; Fraga et al. 2000b; Frysinger et al. 1999; Frysinger and Gaines 2001; Harju and Haglund 2001; Marriott et al. 2000; Prazen et al. 2001; Shellie et al. 2001a]. A GC x GC instrument consists of two chromatographic columns interconnected by a sample modulation interface. The sample modulation interface injects portions of the first column effluent

onto the second column at rapid intervals. The first column is the longer of the two columns, and often has a non-polar stationary phase, while the second column is shorter, often with a polar stationary phase. Thus, GC x GC separations utilize two columns providing complementary separations [Venkatramani et al. 1996]. This leads to a greatly enhanced peak capacity [Giddings 1990], making GC x GC more powerful and less time consuming than traditional 1D GC.

Another popular technique for analyzing complex mixtures is gas chromatography with mass selective detection (GC-MSD). At a basic level "mass spectrometers are analytical instruments that convert neutral molecules into gaseous ions and separate those ions according to the ratio of their mass-to-charge (*m/z*)" [Cotter 1997]. The data from a mass spectrometer is usually displayed as a mass spectrum, which is a plot of relative intensity vs. *m/z*. This information can be used to understand the structure and identity of a substance based upon the differing ratios of ions present in a mass spectrum. By coupling mass spectrometry to gas chromatography, identification of fully separated components is possible using mass spectral database searches or a handful of highly selective ions for specific components. Conclusive identification of analytes without the use of standards is often not possible with commonly used universal detectors such as flame ionization detection (FID). Complex mixtures, however, are usually very difficult to separate using traditional 1D chromatography, resulting in very long run times if separation is possible at all. The most common types of mass spectrometers for GC-MSD have been quadrupole instruments. Although there is a possibility to use a quadrupole MSD as a detector for GC x GC [Debonneville and Chaintreau 2004; Shellie and Marriott 2003; Shellie et al. 2003], the scanning speed of quadrupole MSDs (~1 second for *m/z* 20 to 800) is not fast

enough to use as a detector for the fast second column separations of GC x GC instruments with second column peak widths routinely around 100 ms for applications where a full mass spectrum is required. With the ready availability of fast and powerful computers and advances in electronics, time-of-flight mass spectrometers (TOFMS) have reached the speeds needed to use them as detectors for GC x GC separations.

Some work has been done recently with thermal modulation GC x GC instruments coupled to TOFMS instruments to investigate operating conditions and limits of detection [Dallüge et al. 2002a; Dallüge et al. 2002c; Shellie et al. 2001b; van Deursen et al. 2000]. Not only is this configuration an excellent tool for the identification of components in complex mixtures, but also because it makes a second-order technique a third-order technique, this opens the door for unique chemometric signal deconvolution techniques. Using a chemometric technique such as parallel factor analysis (PARAFAC), it is possible to deconvolute individual components from a group of partially overlapped components using only one sample. With second-order systems, e.g., GC x GC, chemometrics can deconvolute unknowns using a standard and a sample where the analyte of interest varies in concentration between the two data sets. Thus by combining GC x GC with TOFMS detection, a technique that provides three dimensions of selectivity is produced. When combined with a chemometric technique such as PARAFAC, GC x GC-TOFMS may eliminate the need for retention time alignment and most standards for the identification of compounds in a complex mixture even when there is only partial selectivity in both chromatographic separation spaces and the mass spectral dimension.

1.2 Background

1.2.1 Two-Dimensional Comprehensive Gas Chromatography (GC x GC)

Since its debut in 1991 [Liu and Phillips 1991], GC x GC instrumentation has continually evolved. Major differences in instrumentation design stem from differences in the sample modulation interface used to manipulate the first column effluent and re-inject chemical components onto the second column. The most commonly used method is thermal modulation. The first common design consisted of a slotted heater that sweeps over a thick-filmed capillary that connects the columns. The chemical components that have been retained are volatilized and then injected onto the second column when the heater sweeps over the capillary [de Geus et al. 2000; Phillips and Xu 1995; Phillips et al. 1999]. The second method, cryogenic modulation, also involves the manipulation of temperature to make injections on the second column. In this design, a jet of cryogenic CO_2 is applied with a moving trap to a small portion of the capillary between the two columns. The trap moves back and forth over a portion of capillary to accumulate and then inject portions of the first column effluent onto the second column [Kinghorn and Marriott 2000; Marriott and Kinghorn 1997]. Other cryogenic designs use strategically placed and timed jets of cryogenic gas or a combination of heated and cooled jets to eliminate the moving parts [Beens et al. 2001a; Ledford and Billesbach 2000].

The third instrument design employs a diaphragm valve for the sample modulation interface between the two columns. There have been several different configurations of this modulation method. The Synovec lab pioneered the first valve-based design that had both columns and the valve

inside the same GC oven [Bruckner et al. 1998]. This initial valve-based configuration worked for mixtures that could be separated without pushing the temperature beyond 175 °C, the maximum manufacturer-specified operating temperature for the diaphragm valve. More recently a modified design introduced a second independently controlled oven that housed the valve and the second column [Johnson et al. 2002; Prazen et al. 2001]. Independent temperature control results in a more versatile system to optimize separations on both columns. The first column can therefore be subjected to much higher temperatures in this design while maintaining the diaphragm valve at or below 175 °C. Since the effluent from the first column may contain high-boiling components, there is the possible complication of component condensation in the valve or the sample loop. This would be seen as a substantial increase in tailing in the first dimension of late-eluting peaks compared to early eluting peaks. Seeley and co-workers have constructed a valve-based GC x GC in which the valve is mounted between the GC oven and the detector platform. The valve body is maintained at 125 °C while the ports of the valve are pointed towards the oven and follow the oven temperature closely [Seeley et al. 2000]. Both columns are housed inside the main oven and connected to the valve via short segments of deactivated silica capillary. This configuration allows both columns to be taken to higher temperatures and programmed at the same rate, but the underlying potential for excessive heat accumulation in the valve at high oven temperatures and the possibility of high-boiling point components condensing in the valve both exist. Work will be presented in Chapter 2 that develops a valve-based GC x GC design that features a valve placement that extends the workable temperature range to ~250 °C [Sinha et al. 2003a]. The hypothesis being tested here is that by circumventing the temperature limitations of the valve, this

new design will be capable of analyzing complex mixtures that contain both volatile and semi-volatile mixtures.

1.2.2 Time-of-Flight Mass Spectrometry (TOFMS)

TOFMS was first proposed in 1946 by Stephens [Stephens 1946]. The Bendix Corporation later produced the first commercially available instrument in 1957 [Cotter 1994]. Linear TOFMS systems consist of a source region, a drift region held at vacuum (~10^{-7} torr), and a detector. Compounds are ionized at the source, accelerated to the same kinetic energy, and extracted into the drift region. Since all the ion fragments have the same kinetic energy, but different masses, their velocities down the length of the flight tube are inversely proportional to the square root of their masses. Ions with smaller masses have higher velocities, traverse the length of the drift tube in less time, and are therefore detected earlier than the larger ions that have lower velocities. One complicating factor to the relationship between mass and velocity and the corresponding mass resolution is the initial temporal, spatial, and kinetic energy distributions of molecules in the source region. Ions spending a longer time in the source region and those with initial velocities in the direction of the flight tube are accelerated to higher velocities relative to other ions of the same mass in the ion packet. This leads to a distribution of flight times when the ions reach the detector. Using a combination of short ionization pulses and a tightly focused ionization beam helps reduce spatial and temporal distributions. One method for reducing the kinetic energy distribution of the ion packet is to employ a reflectron, a design proposed by Mamyrin and co-workers in 1973 [Mamyrin et al. 1973]. A TOFMS that has a reflectron consists of a source region where the ions are formed, an initial drift region containing ion optics to focus the ion beam, a reflectron (referred to as an ion mirror located where the detector would have been in linear

instruments), a second drift region to achieve separation of the focused ions, and a detector. The ion mirror consists of a series of lens elements with a constant retarding field. Ion focusing occurs because ions with higher velocities penetrate further and ions with lower velocities penetrate less into the ion mirror before being reflected back into the second drift region. The faster ions and the slower ions then converge as soon as they leave the ion mirror. The ion mirror effectively recreates the ion packet as it was at the source with greatly decreased initial energy distributions. The focused ion packet then experiences the second drift region where mass separation occurs due to the time it takes for each group of ions with the same *m/z* to reach the detector.

1.3 Chemometrics

Chemometrics has been described as the art of extracting chemically relevant information from data produced in chemical experiments [Wold 1995]. Although the field of chemometrics has existed since 1972, the introduction of inexpensive, fast computers and the routine interfacing of instruments and computers has accelerated the implementation of chemometric techniques in recent years. By combining 2D separations and chemometrics, run times can be reduced even further and chemical selectivity can be better utilized. Chemometrics substantially extends the information gained from 2D separations by more completely evaluating the analytical data. Three basic categories of chemometrics have been applied to data from 2D separations: peak deconvolution, multivariate calibration and pattern recognition [Synovec et al. 2003b]. In general, chemometrics is applied to 2D separation data to further realize the selectivity of the data and to glean more useful information.

In order to have a basis for better understanding the potential benefit of combing 2D separations with chemometrics, it is important to apply useful figures-of-merit to evaluate the performance of 2D separations that are then subjected to chemometric analysis. Chromatographic resolution, R_s, is a good metric for judging the separation of two peaks, but often it does not suffice when analyzing low-resolution 2D separations. R_s describes the separation of a peak overlapped with a single interfering peak, but when there are multiple interfering peaks R_s cannot fully describe the information. This shortcoming is complicated further by multiple dimensions of separation, where the net R_s of peaks is a combination of the R_s on each dimension.

A useful metric to address the shortcomings of chromatographic resolution is called "multivariate selectivity" [Booksh and Kowalski 1994; Faber et al. 1997; Lorber et al. 1997]. Multivariate selectivity measures the degree of overlap of signals from different chemical components. For 2D separation analysis, multivariate selectivity can describe the information content of peaks overlapped with multiple interfering peaks and can be extended to describe data from 2D separations hyphenated with spectrometric instruments. Recent experimental data has shown that multivariate selectivity is strongly correlated with the quality of quantitative results when chemometric deconvolution techniques are used to quantify peaks that are overlapped with multiple interfering peaks. The relationship between 2D chromatographic resolution (2D-R_s) and multivariate selectivity is shown in Figure 1.1. 2D-R_s has been defined as the square root of the sum of the squared R_s on the two dimensions [Murphy et al. 1998]. For example, having a 1D-R_s of 0.7 on each dimension corresponds to a 2D-R_s of 1.0. Multivariate selectivity ranges between zero (complete overlap) and unity (no overlap). Multivariate selectivity increases quickly with resolution until the 2D-R_s is

about 0.75 (Figure 1.1). Above a 2D-R_s of 0.75, since the two peaks are nearly resolved, the multivariate selectivity approaches the optimum value and there is little additional analytical information to be gained by resolving them further, e.g., using a larger 2D-R_s is excessive. Implementing 2D separations and univariate detection (e.g., FID with GC x GC) with chemometrics for peak deconvolution and quantification provides good results for peaks with a 2D-R_s ~ 0.30, corresponding to a multivariate selectivity of ~ 0.35 [Fraga et al. 2001a]. If a 2D-R_s of 1.5 is required for analysis without chemometrics, reducing the R_s requirement to 0.3 translates into a separation run-time savings of a factor of 25. This is a substantial benefit.

1.4 Peak Deconvolution

One application of chemometrics is deconvolution, or the mathematical resolution, of overlapped peaks. The generalized rank annihilation method (GRAM) is a chemometric technique that uses the bilinear structure of 2D separations to both deconvolve peaks and to compare a peak's magnitude to that of a standard for quantitation. GRAM has been successfully applied to GC x GC, to quantify partly resolved peaks, e.g., for the analysis of selected overlapped components in modified white gasoline [Bruckner et al. 1998] and jet fuel [Fraga et al. 2000b]. To determine the applicability of GRAM to other comprehensive 2D methods, Monte Carlo simulations have been applied to determine the conditions where the use of GRAM provides a successful analysis of unresolved peaks [Fraga et al. 2001a]. A wide range of experimental conditions and performance criteria typical to 2D separation methods were modeled: analyte/interference peak height ratio, first- and second-dimension R_s, signal-to-noise ratio, injection volume reproducibility, and run-to-run retention time reproducibility. According to the results of these simulations, peak deconvolution techniques like GRAM should greatly

expand the number of analyzable peaks for all types of comprehensive 2D separations [Fraga et al. 2001a].

Consistent run-to-run retention times are crucial to the application of chemometrics; hence, research into effective data alignment algorithms is imperative. The more precise the run-to-run retention times, the more powerful the chemometric analysis becomes. In this regard, an objective retention time standardization technique for 2D separation data significantly improves quantitation in a GRAM analysis [Prazen et al. 1999b]. This alignment technique has been expanded to correct both dimensions using interpolated retention time shifts to enhance precision [Fraga et al. 2001b]. These algorithms are rank-based and thus require an estimation of the number of independent chemical components present (rank) in the region (submatrix) of the peaks being deconvolved. Other alignment techniques that have been developed for 1D separations should be extendable to 2D separation [Johnson et al. 2003].

Although the application of multivariate curve resolution techniques to 2D separations has yet to be realized, the success of these methods for the peak deconvolution of data from separation techniques hyphenated with spectroscopic techniques indicates that they will be powerful tools for the analysis of 2D separation data [Johnson et al. 1999].

1.5 Multivariate Calibration

Trilinear partial least squares (Tri-PLS) is an extension of the popular chemometric technique partial least squares (PLS) to 2D data. Tri-PLS and other multivariate calibration techniques do not require peak identification while taking advantage of the redundant measurements of each analyte peak within the 2D separation. GC x GC with tri-PLS analysis was demonstrated on naphtha samples in a study geared toward monitoring and control of

complex process streams [Prazen et al. 2001]. High-speed GC x GC analysis with a short column, high linear flow velocity, and partial chromatographic resolution was followed by tri-PLS analysis. This resulted in separations 16 times faster than the traditional 1D GC method. Tri-PLS was used to predict the aromatic and naphthalene (cycloalkane) portion of the naphtha samples and achieved acceptable quantitative precision. This is evidence that a great increase in analysis speed can be realized with the combination of 2D chromatographic methods and chemometrics, which can increase throughput and productivity.

1.6 Pattern Recognition

Recently, three-way principal component analysis (PCA) has been used to analyze 2D-PAGE data from proteomic samples to identify classes of the samples present [Marengo et al. 2003]. Two different types of sample sets were analyzed: a set of rat sera to distinguish which had been subjected to nicotine and a set of human lymph-nodes to distinguish between a class of healthy lymph-nodes and a class affected by a non-Hodgkin's lymphoma. Three-way PCA successfully distinguished between the two classes in each case and was able to elucidate regions in the 2D maps responsible for the class differentiation.

Pattern recognition capabilities have also been demonstrated with GC x GC data. ANOVA-based feature selection was used prior to PCA to elucidate and select features pertinent to class distinction for jet fuel mixtures [Johnson and Synovec 2002]. After the feature selection, PCA was able to distinguish a 1% volumetric compositional difference in mixtures of two jet fuel types. In addition, three types of jet fuel (and mixtures of two or three of the fuels) from various locations throughout the United States were analyzed to demonstrate that feature selection and fuel classification was possible with

this technique in the presence of geographical variations. In this study ANOVA-based feature selection was demonstrated to be a useful tool to enhance chemical selectivity and the classification power of 2D separations when combined with PCA.

1.7 Hyphenated Comprehensive Two-Dimensional Separations

Some of the most recent advancements for 2D separations and chemometrics are with the addition of multivariate spectrometric detection, such as TOFMS or multi-wavelength absorbance detection [Dallüge et al. 2003]. By adding a spectrometric detector to a 2D separation, an instrument that is inherently trilinear is created, producing a third order data structure for each sample. Figure 1.2 illustrates that the total signal in trilinear data is equal to the sum of the outer products of the signal in each dimension for each chemical component in the sample plus an error (or noise) matrix. This portends that only one data set is needed to perform peak deconvolution by chemometric techniques that rely on the third order data structure without peak shape assumptions. This is termed the third-order advantage [Booksh and Kowalski 1994]. This also eliminates the dependency on retention time alignment for peak deconvolution. The construction and evaluation of the first valve-based GC x GC-TOFMS is presented in Chapter 3. The hypothesis being tested in this chapter is that a valve-based GC x GC instrument can be coupled to a TOFMS providing highly selective detection in light of complex flow regimes in valve-based instruments.

The work presented in Chapters 4 – 7 covers recent progress in combining GC x GC-TOFMS data with chemometric analysis. In Chapter 4, the hypothesis that GC x GC-TOFMS data is compatible with multivariate

chemometric techniques is tested by deconvoluting overlapped peaks in a complex environmental sample (#1) using PARAFAC [Sinha et al. 2004a], implementing the third-order advantage. Chlorobenzene is successfully deconvoluted, identified and quantified in the environmental sample (#1) in the presence of substantial background interference and an overlapping compound. Chapter 5 investigates of the practical limitations of combining GC x GC-TOFMS with PARAFAC for the analysis of very closely related isomers, a significant chemical analysis challenge in complex mixtures. The hypothesis being tested here is that PARAFAC can successfully deconvolute signals with no purely selective mass channel, but some selectivity in each dimension. Studies were performed to gauge the quality of PARAFAC deconvolution as a function of chromatographic and mass spectral selectivity in GC x GC-TOFMS data [Sinha et al. 2004b]. In Chapter 6, a cursory, yet insightful comparison of deconvolution techniques applicable to GC x GC-TOFMS data is presented. One analyte of interest in a complex region of an environmental sample (#2) is deconvoluted using PARAFAC, AMDIS, and the LECO ChromaTOF software deconvolution program. The hypothesis being tested here is that a chemometric technique that utilizes the third-order data structure (e.g., PARAFAC) offers unique benefits over those that do not. Chapter 7 tests the hypothesis that a novel algorithm based on mass spectral similarity can successfully locate analytes of interest in a GC x GC-TOFMS data set in an automated fashion. This new chemometric tool is used to elucidate the locations of several analytes of interest in a derivatized human infant urine sample.

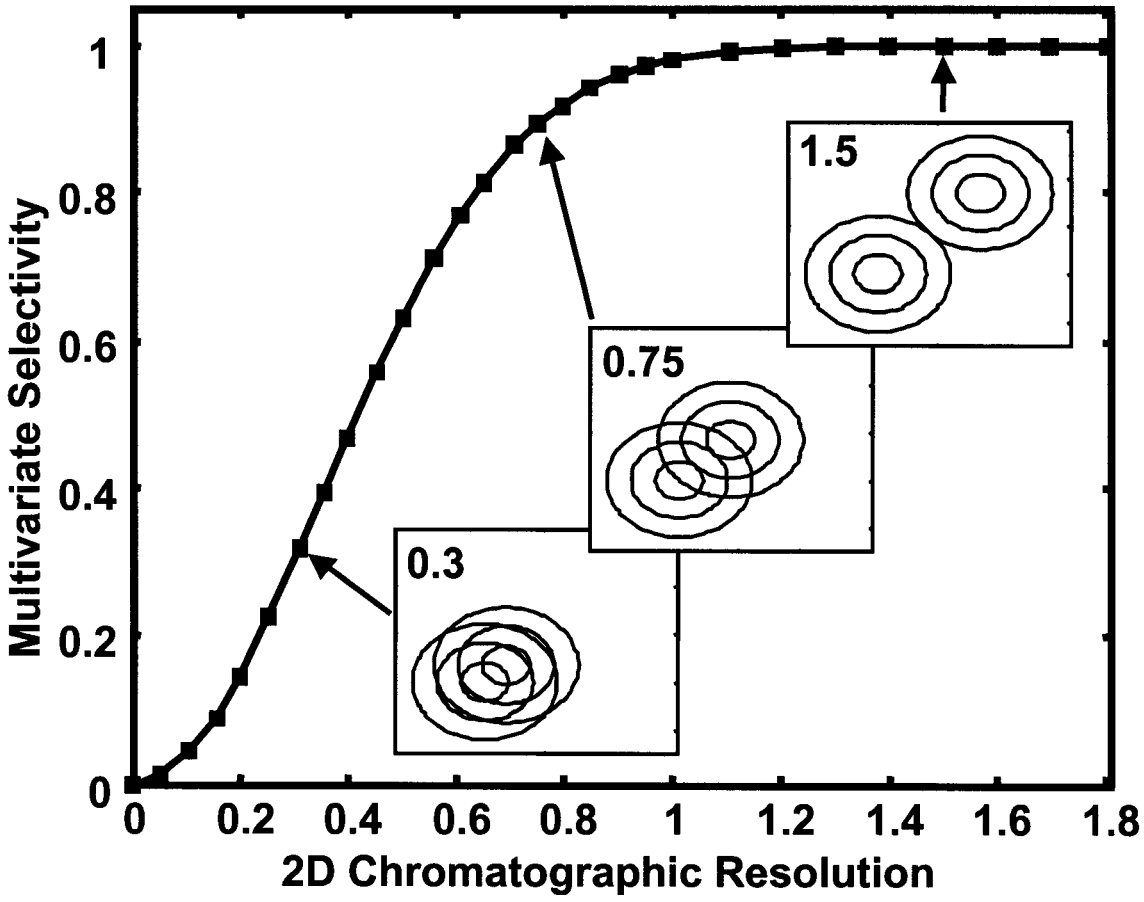


Figure 1.1. Graph of multivariate selectivity as a function of 2D chromatographic resolution. Three examples of the 2D-R_s of two peaks along a diagonal in a 2D separation are shown.

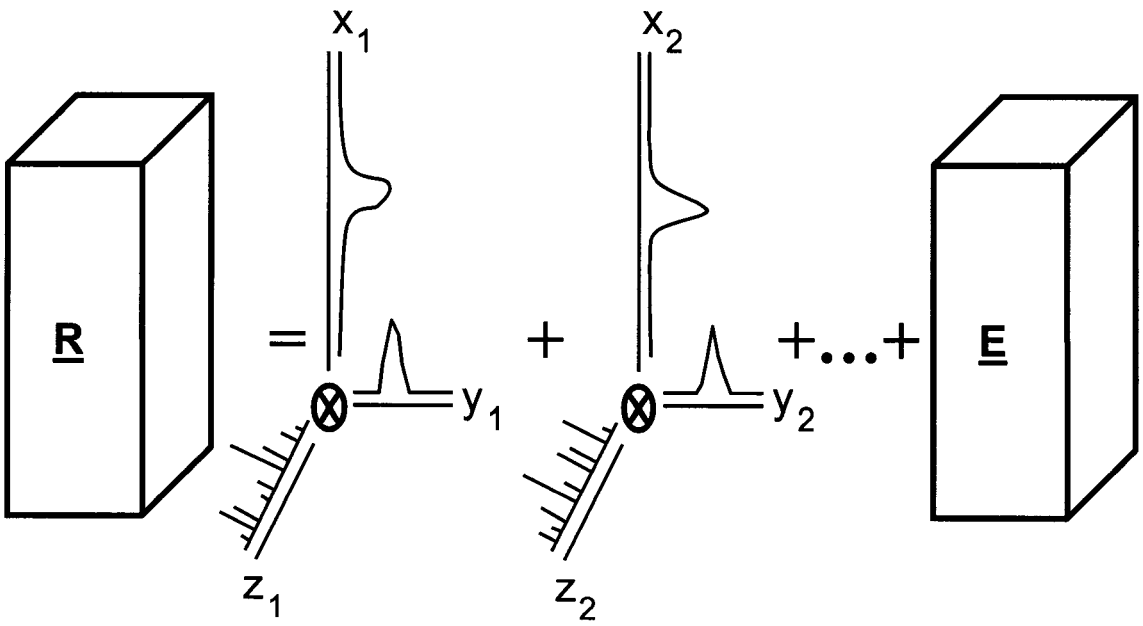


Figure 1.2. Illustration of the trilinear data structure of GC x GC-TOFMS data. For the instrument response $\underline{\mathbf{R}}$ there are distinctive profiles in both chromatographic dimensions ($\mathbf{x}_n$ and $\mathbf{y}_n$) and a unique spectrum ($\mathbf{z}_n$) for each constituent in a data matrix. The error (e.g., noise) is denoted $\underline{\mathbf{E}}$. This is mathematically described as $\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}}$.

CHAPTER 2. Comprehensive Two-Dimensional Gas Chromatography (GC x GC) of Volatile and Semi-Volatile Components using a Diaphragm Valve-Based Instrument*

2.1 Introduction

Two-dimensional comprehensive gas chromatography (GC x GC) is a powerful technique that is well suited to analyzing complex chemical mixtures [Beens et al. 2001b; Bertsch 2000; de Geus et al. 1998; de Geus et al. 2001; Dimandja et al. 2000; Fraga et al. 2000b; Frysinger et al. 1999; Frysinger and Gaines 2001; Harju and Haglund 2001; Marriott et al. 2000; Prazen et al. 2001; Shellie et al. 2001a]. A GC x GC instrument consists of two chromatographic columns interconnected by a sample modulation interface. The sample modulation interface injects portions of the column 1 effluent onto column 2 at rapid intervals. Column 1 is the longer of the two columns, and often has a non-polar stationary phase, while column 2 is shorter, often with a polar stationary phase. Thus, GC x GC separations utilize two columns providing complementary separations [Venkatramani et al. 1996]. This leads to a greatly enhanced peak capacity [Giddings 1990], making GC x GC more powerful and less time consuming than traditional one-dimensional GC.

Since its debut in 1991 [Liu and Phillips 1991], GC x GC instrumentation has continually evolved. Major differences in instrumentation design stem from differences in the sample modulation interface used to manipulate the column 1 effluent and re-inject chemical

---

* Large portions of this Chapter have been previously published [Sinha et al. 2003a].

components onto column 2. The most commonly used method is thermal modulation. The initial design consists of a slotted heater that sweeps over a thick-filmed capillary that connects the columns. The chemical components that have been retained are volatilized and then injected onto column 2 when the heater sweeps over the capillary [de Geus et al. 2000; Phillips and Xu 1995; Phillips et al. 1999]. The second method, cryogenic modulation, also involves the manipulation of temperature to make injections on column 2. In this design, a jet of cryogenic CO_2 is applied with a moving trap to a small portion of the capillary between the two columns. The trap moves back and forth over a portion of capillary to accumulate and then inject portions of the column 1 effluent onto column 2 [Kinghorn and Marriott 2000; Marriott and Kinghorn 1997]. Other designs use strategically placed and timed jets of cryogenic gas or a combination of heated and cooled jets to eliminate the moving parts [Beens et al. 2001a; Ledford and Billesbach 2000].

The third instrument design employs a diaphragm valve for the sample modulation interface between the two columns. There have been several different configurations of this modulation method. The first design had both columns and the valve inside the same GC oven [Bruckner et al. 1998]. This initial configuration worked for mixtures that could be separated without pushing the temperature beyond 175 °C, the maximum manufacturer-specified operating temperature for the diaphragm valve. More recently a modified design introduced a second independently controlled oven that housed the valve and column 2 [Johnson et al. 2002; Prazen et al. 2001]. Independent temperature control results in a more versatile system to optimize separations on both columns. Column 1 can therefore be subjected to much higher temperatures in this design while maintaining the diaphragm valve at or below 175 °C. Since the effluent from column 1 may contain high-

boiling components, there is the possible complication of component condensation in the valve or the sample loop. This would be seen as a substantial increase in tailing in the first dimension of late-eluting peaks compared to early eluting peaks. Seeley and co-workers have constructed a valve-based GC x GC in which the valve is mounted between the GC oven and the detector platform. The valve body is maintained at 125 °C while the ports of the valve are pointed towards the oven and follow the oven temperature closely [Seeley et al. 2000]. Both columns are housed inside the main oven and connected to the valve via short segments of deactivated silica capillary. This configuration allows both columns to be taken to higher temperatures and programmed at the same rate, but the underlying potential for excessive heat accumulation in the valve at high oven temperatures and the possibility of high-boiling point components condensing in the valve both exist.

In the current investigation a new valve-based GC x GC configuration is reported that manipulates the valve placement to circumvent the temperature specified limitations of the diaphragm valve. The most commonly used diaphragm valve has a temperature limit of 175 °C due to three O-rings in the interior of the valve. Operating the valve beyond the manufacturer's specifications can result in unsatisfactory valve performance over time. Fortunately, the O-rings are located at the opposite end of the valve relative to the position of the valve diaphragm, valve ports and sample loop. Thus, by maintaining the portion of the valve that is in contact with sample components at the same temperature as the columns while keeping the portion of the valve containing the O-rings below 175 °C, the temperature limitations of the valve can be overcome. In this new instrument configuration, the valve is mounted such that the valve diaphragm, valve

ports and sample loop are inside the GC oven and the rest of the valve is outside the oven, freely exposed to room air. This new configuration allows a much larger workable temperature range for valve-based GC x GC instruments than previous configurations. A separation of a test mixture containing volatile and semi-volatile components is demonstrated, with boiling points ranging from 65 °C to 270 °C. Data quality was evaluated, such as retention time precision and achievement of a bilinear data structure, both important criteria for chemometric data analysis. Temperature studies were conducted in order to ascertain that the sample loop, valve ports and valve diaphragm were held at or sufficiently near the oven temperature while the valve O-rings were kept well below the maximum specified temperature of 175 °C.

2.2 Experimental

In the valve-based comprehensive GC x GC configuration modified for high temperature operation (Figure 2.1), both columns were housed inside the same oven. Sample traveled through column 1 and into a 1.3 μl sample loop on a high-speed six-port micro diaphragm valve (VICI, Valco Instruments Co. Inc. Houston, TX, USA). The effluent in the sample loop from column 1 was injected periodically onto column 2 and then detected with a flame ionization detector (FID). The column 1 effluent was vented during each injection onto column 2. The valve was mounted partially inside the oven so that the sample loop and the diaphragm were inside the oven and the temperature sensitive O-rings were outside the oven (Figure 2.2). A Varian 3600 CX gas chromatograph (Varian Inc., Palo Alto, CA, USA) was equipped with an HP 7673A automatic sampler (Hewlett Packard, San Fernando, CA USA) and modified by cutting a 1 3/8" hole in the top of the oven to allow for the new

valve placement. A tight fit between the valve and oven wall was achieved so the oven performance was not compromised. An in-house LabVIEW 6i (National Instruments, Austin, TX, USA) program and a data acquisition board (model AT-MIO-16XE - 50, National Instruments) were used to control the valve actuation and to collect the data at a rate of 20,000 points per second boxcar averaged to 250 points per second. All data processing was accomplished with Matlab 6.0 R12 (The Mathworks, Natick, MA, USA).

Temperature studies on the valve were performed using four type-K thermocouples (chromel - alumel). The precision and accuracy of the thermocouples was determined by recording the temperature of boiling water for five minutes. The thermocouples agreed to the standard (100 °C) and to each other to within less than 0.4%. Three of the thermocouples were attached to the valve using high-temperature Kapton® polyimide tape (ESD Systems.com, Marlboro, MA, USA) (Figure 2.2). The Top-Valve thermocouple was mounted on the outside of the valve nearest the location of the O-rings. The Mid-Valve thermocouple was mounted on the surface of the valve just above the top of the oven. The Sample Loop thermocouple was mounted on the sample loop inside the oven. The fourth thermocouple ("Oven") was used to record the oven temperature and was located approximately halfway down near the door of the oven. Temperature data were recorded using a LabVIEW program that uses an eighth-order polynomial and the built in cold-finger compensator on the National Instruments DAQ board to convert the voltage measurement of the thermocouple to a temperature in degrees Celsius.

A test solution of 26 components of homologous series of n-alkanes (C6 - C15), 2-ketones (C4 - C9), and n-alcohols (C3 - C11) was mixed from PolyStandard™ Kits (AccuStandard, Inc., New Haven, CT, USA). An equal volume of each component was added to the solution resulting in a ~3.8% v/v

concentration for each test compound. Boiling points for all components are listed in Table 2.1 and range from 65 °C to 270 °C.

A GC x GC separation was performed on the 26-component test mixture to demonstrate the separation characteristics of the new configuration. Column 1 of the GC x GC was a 12-m 200-μm i.d. capillary column with a 0.33-μm poly(dimethylsiloxane) film (SPB-1; Supelco, Bellfonte, PA, USA). The second column was a 0.5-m 100-μm i.d capillary column with a 0.2-μm poly(ethylene glycol) film (CP-Wax 52 CB; Chrompack, Varian, Palo Alto, CA, USA). The valve was actuated with one-second cycle times with a 20 ms injection pulse width. Helium was used as the carrier gas. Both columns were operated under constant head pressure. Column 1 had a head pressure of 4 psi providing an initial flow of 0.3 ml/min and column 2 had a head pressure of 20 psi resulting in an initial flow of 3.55 ml/min. The oven was held at 40 °C for 1 minute and then programmed to ramp to 250 °C at 15 °C/min. A 1-μl injection of the 26-component mixture was split 100:1.

2.3 Results and Discussion

The temperature effects on the valve during a typical chromatographic run were studied by recording a temperature reading every second from the four type-K thermocouples attached to the valve (Figure 2.2). Representative data are reported herein. The oven was programmed to ramp four times from 50 °C to 250 °C at a rate of 15 °C/min and the results of the fourth ramp are shown in Figure 2.3A. The other temperature program runs were essentially identical but not shown for brevity. The data indicate that the temperature of the Sample Loop sensor position closely follows the temperature of the Oven sensor position. The portions of the valve outside the oven, as monitored at the Top-Valve and Mid-Valve sensor positions, are significantly lower in

temperature throughout the entire temperature program. The sample loop reached 247 °C at the end of the temperature program, deviating from the oven set point of 250 °C by only 1.2%. The O-rings, situated inside the valve and between Top-Valve and Mid-Valve, reached a temperature between 79.5 °C (Top-Valve) and 82.7 °C (Mid-Valve) by the end of the temperature program. The temperature of the O-rings were well below the maximum operating temperature of 175 °C provided the temperature on the surface of the valve is not substantially different to the internal temperature of the valve near the O-rings. Since the position of the Mid-Valve temperature sensor is substantially closer to the oven then any of the O-rings, and the Mid-Valve temperature reading was close to the Top-Valve temperature reading, it is reasonable to presume the O-rings are indeed in an environment well below 175 °C. While it is difficult to fully prove this condition, one must rely upon the performance of the system over extended time of testing, as was done in this work.

The potential of heat accumulation in the valve was studied by holding the GC oven at 240 °C for a period of 3 hours, recording temperature data every second and then boxcar averaging every five minutes. Figure 2.3B contains the temperature profiles for the four thermocouples after a steady-state temperature in the Top-Valve and Mid-Valve positions was reached. For over two hours the temperatures on the valve recorded inside and outside the oven were relatively constant. The area containing the O-rings had an average temperature between 129.2 ± 0.6 °C (Top-Valve) and 138.3 ± 0.5 °C (Mid-Valve). This was well below the limit of 175 °C although higher than the maximum temperatures recorded during the temperature program study. This suggests that there is a potential for heat to build up in the valve over time, but not enough to exceed acceptable operating temperatures under

normal conditions. Inside the oven, the sample loop reached a steady-state temperature of 241.0 ± 0.1 °C (Sample Loop) while the oven had a steady-state temperature of 248.1 ± 0.1 °C (Oven). The sample loop temperature only differs from the set point of 240 °C by 0.5% while the measured oven temperature 248.3 °C differs by 3.5% from the oven set point. This discrepancy is due to the vertical temperature variations within the GC oven itself. The built-in thermocouple used to regulate oven temperature is located in the upper-rear right corner at a similar vertical position as the thermocouple on the sample loop, while the "Oven" thermocouple was located in the vertical middle near the front of the GC oven.

After six months of use, the valve in this new configuration has not suffered any noticeable deterioration in performance. The system has also been temperature programmed to 265 °C, a 90 °C increase from the previous maximum operating temperature of 175 °C. At these high temperatures the stability of many stationary phases (e.g., polar) is an issue. Most commonly used polar phases (phenyl, cyano, and PEG) have maximum temperature limits of around 250 °C. Thus, the temperature limits of the polar stationary phases may be the most prominent limitation to further increasing the maximum temperature of valve-based GC x GC instruments.

A 26-component test mixture of alkanes, ketones, and alcohols was utilized to investigate the characteristics of the new system configuration. Table 2.1 lists the individual components of the mixture and their boiling points. Figure 2.4A depicts a three dimensional (3D) surface plot of the separation for this mixture. The more highly retained large alcohols such as 1-decanol (A-9) and 1-undecanol (A-10) are the least prominent peaks in relation to other components in the mixture. This is caused by peak broadening due to a combination of decreasing flow rate on column 2 with increasing

temperature and a relatively thick stationary phase that highly retains alcohols. Peak widths at 10% of the peak height for representative analytes range from 6.58 s to 9.69 s on column 1 and 0.101 s to 0.211 s on column 2 (Table 2.2). Figure 2.4B depicts a contour plot of the same separation of this mixture with alcohols labeled A1-A10, ketones labeled B1-B6, and alkanes labeled C1-C10. By temperature programming both columns, it is possible to produce a well-tuned separation [Phillips and Beens 1999]. In this type of separation, the homologous series elute in different near-horizontal bands across the two-dimensional separation space dependent upon their retention on column 2. In this way, the separation space is used more uniformly. Note that an excellent separation is achieved for the test mixture containing components ranging from volatile species (65 °C lowest boiling point) to semi-volatile species (270 °C being the highest boiling point).

The retention time reproducibility of the new configuration was studied. Representative data were reported, herein, by comparing four replicate sets of data. Table 2.3 summarizes the mean retention time and standard deviation for four test compounds selected to represent and span the two-dimensional separation space: 1-butanol (A3), 2-heptanone (B4), 1-nonanol (A8), and n-tridecane (C8). The sub-matrix signal of 1-nonanol, A8, is shown in Figure 2.5A. The sub-matrix is organized such that each second column profile is a column of data in the matrix. The run-to-run retention time variability for each compound on both the column 1 time axis and the column 2 time axis was determined. For each run of the test mixture, the column 1 peak profile for each test compound was obtained by summing all of the data from each column 2 run onto the column 1 time axis using the sub-matrix for each isolated signal. The first moment of the resulting peak profile was determined, providing the retention time. The mean retention time and

standard deviation for each test compound was determined using the four replicate runs. The run-to-run retention time variability for each test compound on the column 2 time axis was similarly determined. The run-to-run retention time variability on column 1 was less than 0.1% for each of the four test compounds of interest. The run-to-run retention time variability on column 2 was greatest for highly retained compounds, 0.7% for 1-nonanol (A8), and lower for less retained compounds, 0.1% for 2-heptanone (B4). The within-run retention time variability on column 2 was analyzed by calculating the first moment of each column of data in each test compound sub-matrix. The mean and standard deviations of the retention times for each data column comprising the peak were calculated for each of the four replicate runs. The average of these four means and standard deviations is reported as the within-run variability of column 2 retention times. The within-run standard deviations are greater than run-to-run variations; however, the largest of the variations is still only 1.4% indicating that, in general, the retention time reproducibility is quite good.

One test to determine whether the within-run retention time variation has detrimental effects on the data structure is to employ a multivariate data analysis technique that requires bilinear data. The generalized rank annihilation method (GRAM) has been shown to quantify and deconvolute bilinear data using a calibration data set containing the chemical species of interest [Bruckner et al. 1998; Sánchez and Kowalski 1986; Wilson et al. 1989]. GRAM was applied to the data for the test compound with the worst within-run retention time variation in Table 2.3, i.e., 1-nonanol (A8), to determine whether or not data from this high-temperature configuration is well suited for multivariate data analysis. A sub-matrix of the 1-nonanol data was isolated from each of the four replicate data sets (Figure 2.5A). Typically

GRAM requires the calibration data set and the sample data sets to have differing concentrations; however, with the unique case of having only one component in the matrix, using one of the replicate runs as the standard data set and the other three replicate runs as samples is feasible. For example, using the standard and one of the samples, application of GRAM resulted in the reconstructed peak shown in Figure 2.5B. The GRAM result in Figure 2.5B is essentially identical to the raw data depicted in Figure 2.5A. The three trials gave reproducibly good results as shown in the summed peak profiles for each dimension (Figure 2.5C and D). Therefore, the data from the high-temperature valve-based GC x GC configuration is appropriate for multivariate data analysis.

The shape of the column 1 peak profile of 1-nonanol is also significant because there is no substantial tailing in the column 1 time dimension (Figure 2.5C). If the valve and sample loop were not accurately tracking the oven temperature, a compound with a high boiling point such as 1-nonanol would encounter a "cold spot" in the sample loop upon leaving column 1. The resulting condensation would leave behind some of the compound during each injection onto column 2. This condition would manifest itself as tailing on the peaks in the column 1 time dimension. One method for assessing peak tailing, i.e., peak asymmetry, is to graphically measure empirical asymmetry factors (B/A) [Foley and Dorsey 1983]. To obtain the B/A ratio for a given peak, one must measure the peak width (A) at 10% peak height for the leading portion of the peak, and divide that into the peak width (B) at 10% peak height for the trailing portion of the peak. Both A and B are measured from the time at the peak maximum such that A + B is equal to the total peak width at 10% peak height. The asymmetry factors (B/A) for representative analytes are listed in Table 2.2. A value of exactly 1 for B/A indicates a perfectly

symmetric (Gaussian) peak where a value greater than 1 indicates peak tailing. Column 1 peak profiles have B/A values ranging from 1.09 ± 0.08 to 1.2 ± 0.2 indicating nearly Gaussian peaks within the error of the measurement. Thus, the lack of significant peak tailing in the column 1 time dimension for high-boiling point compounds is an indication that the sample loop and portions of the valve that interact with the sample are closely following the oven temperature and that condensation in the valve is not a significant effect. Column 2 peak profiles have B/A values ranging from 1.14 ± 0.04 to 1.81 ± 0.09 indicating some moderate asymmetry. The asymmetry factor is highest for 1-nonanol, a high-boiling polar compound that is highly retained on column 2. This is likely due to a combination of decreasing flow rate on column 2 with increasing temperature and a thick stationary phase that highly retains alcohols.

2.4 Conclusions

A novel configuration for high-temperature diaphragm valve-based GC x GC has been proposed and tested. Mounting the valve only partially inside the oven has circumvented the valve temperature limit of 175 °C. The detected peak signal profiles for high-boiling compounds such as 1-nonanol (b. p. = 215 °C) indicate that condensation in the valve is not observed. The resulting data obtained with this system conforms to requirements necessary to utilize chemometric data analysis techniques such as GRAM. The workable temperature range has been extended to at least 250 °C and is now applicable to semi-volatile sample components. The new valve configuration shows promise for analyzing complex mixtures with a wider range of boiling points than was previously possible with diaphragm valve-based GC x GC.

Table 2.1. Boiling points for compounds in a 26-component mixture. Test mixture includes alkanes, ketones, and alcohols. Each component is ~3.8 v/v% of the mixture. *Boiling point of 1-undecanol calculated from linear fit of the boiling points of alcohol homologous series. Lit value = 146 °C at 30 mm Hg.

Label	Component	Boiling Point, °C
A1	ethanol	78
A2	1-propanol	97
A3	1-butanol	118
A4	1-pentanol	137
A5	1-hexanol	157
A6	1-heptanol	176
A7	1-octanol	196
A8	1-nonanol	215
A9	1-decanol	231
A10	1-undecanol	249*
B1	2-butanone	76
B2	2-pentanone	98
B3	2-hexanone	124
B4	2-heptanone	147
B5	2-octanone	173
B6	2-nonanone	192
C1	n-hexane	65
C2	n-heptane	98
C3	n-octane	126
C4	n-nonane	151
C5	n-decane	174
C6	n-undecane	192
C7	n-dodecane	216
C8	n-tridecane	232
C9	n-tetradecane	253
C10	n-pentadecane	270

Table 2.2. Peak width data for representative analytes. Column 1 and Column 2 mean peak widths at 10% peak height and mean empirical asymmetry factors (B/A) and standard deviations for 1-butanol (A3), 2-heptanone (B4), 1-nonanol (A8) and n-tridecane (C8) determined following the method described in the text [Foley and Dorsey 1983]. Experimental Conditions: Column 1: 12-m 200-μm i.d. capillary column with a 0.33-μm poly(dimethylsiloxane) film with 4 psi head pressure; Column 2: 0.5-m 100-μm i.d capillary column with a 0.2-μm poly(ethylene glycol) film with 20 psi head pressure; Valve actuation: one second cycles, 20 ms injection widths; Oven Program: held at 40 °C for 1 minute then programmed to reach 250 °C at 15 °C/min; Carrier Gas: Helium; Injection: 1 μl split 100:1.

Test Compound	Col. 1 mean widths at 10% peak height and standard deviation (s)	Col. 1 mean empirical asymmetry factor (B/A) and standard deviation	Col. 2 mean widths at 10% peak height standard deviation (s)	Col. 2 mean empirical asymmetry factor (B/A) and standard deviation
1-butanol (A3)	7.44 ± 0.27	1.2 ± 0.2	0.111 ± 0.003	1.18 ± 0.04
2-heptanone (B4)	6.58 ± 0.04	1.1 ± 0.2	0.101 ± 0.002	1.14 ± 0.04
1-nonanol (A8)	9.69 ± 0.20	1.2 ± 0.2	0.211 ± 0.007	1.81 ± 0.09
n-tridecane (C8)	9.54 ± 0.04	1.09 ± 0.08	0.101 ± 0.001	1.2 ± 0.1

Table 2.3. Retention time reproducibility. Column 1 mean retention times and Column 2 mean retention times with run-to-run standard deviations, and Column 2 mean retention times with mean within-run standard deviations for 1-butanol (A3), 2-heptanone (B4), 1-nonanol (A8) and n-tridecane (C8). Experimental Conditions: Same as for **Table 2.2.**

Test Compound	Col. 1 mean R_t and run-to-run standard deviation (min)	Col. 2 mean R_t and run-to-run standard deviation (s)	Col. 2 mean R_t and mean within-run standard deviation (s)
1-butanol (A3)	3.994 ± 0.005	0.517 ± 0.003	0.518 ± 0.007
2-heptanone (B4)	6.678 ± 0.004	0.3887 ± 0.0004	0.389 ± 0.005
1-nonanol (A8)	10.117 ± 0.004	0.688 ± 0.005	0.69 ± 0.01
n-tridecane (C8)	11.570 ± 0.003	0.350 ± 0.002	0.351 ± 0.002

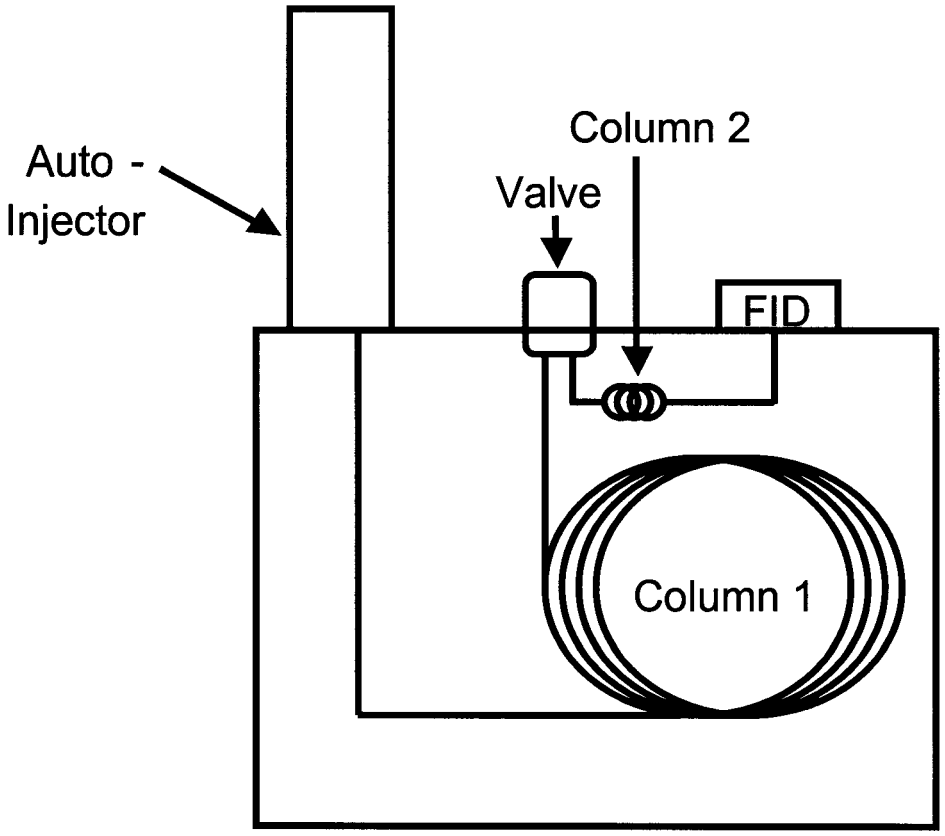


Figure 2.1. GC x GC configuration with dual-column temperature program capabilities. Sample is injected with an auto-injector onto column 1 (AT-1; 12 m x 200 μm x 0.33 μm film thickness). Portions of the sample are injected onto column 2 (CP Wax-52; 0.5 m x 100 μm x 0.2 μm film thickness) from a sample loop on a 6-port mini-diaphragm valve. Effluent from column 2 is detected with a flame ionization detector (FID). A LabVIEW program, written in-house, is used to control the valve and collect data.

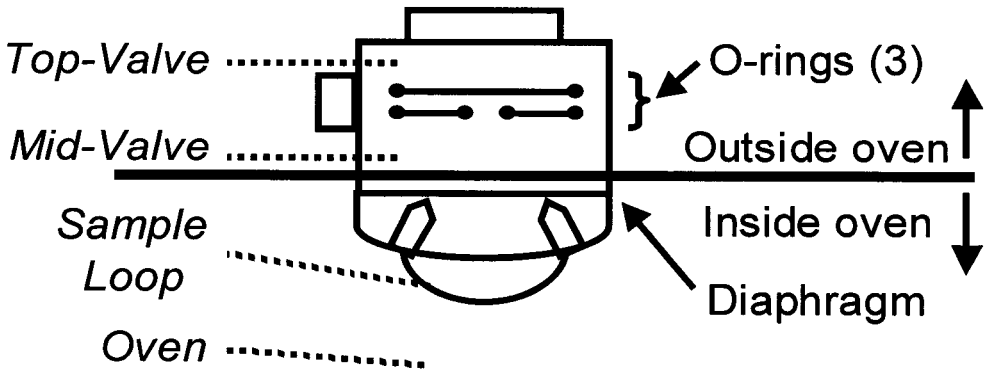


Figure 2.2. Cross-sectional view of the 6-port Mini-Diaphragm Valve (VICI Valco instruments). The valve is mounted with the sample loop and diaphragm inside the oven and the temperature sensitive O-rings outside the oven. Placement of type-K thermocouple sensors at three locations (Top-Valve, Mid-Valve and Sample Loop), attached to the surface of the valve using high temperature polyimide tape. Another thermocouple sensor (Oven) is located in the middle, front of the GC oven.

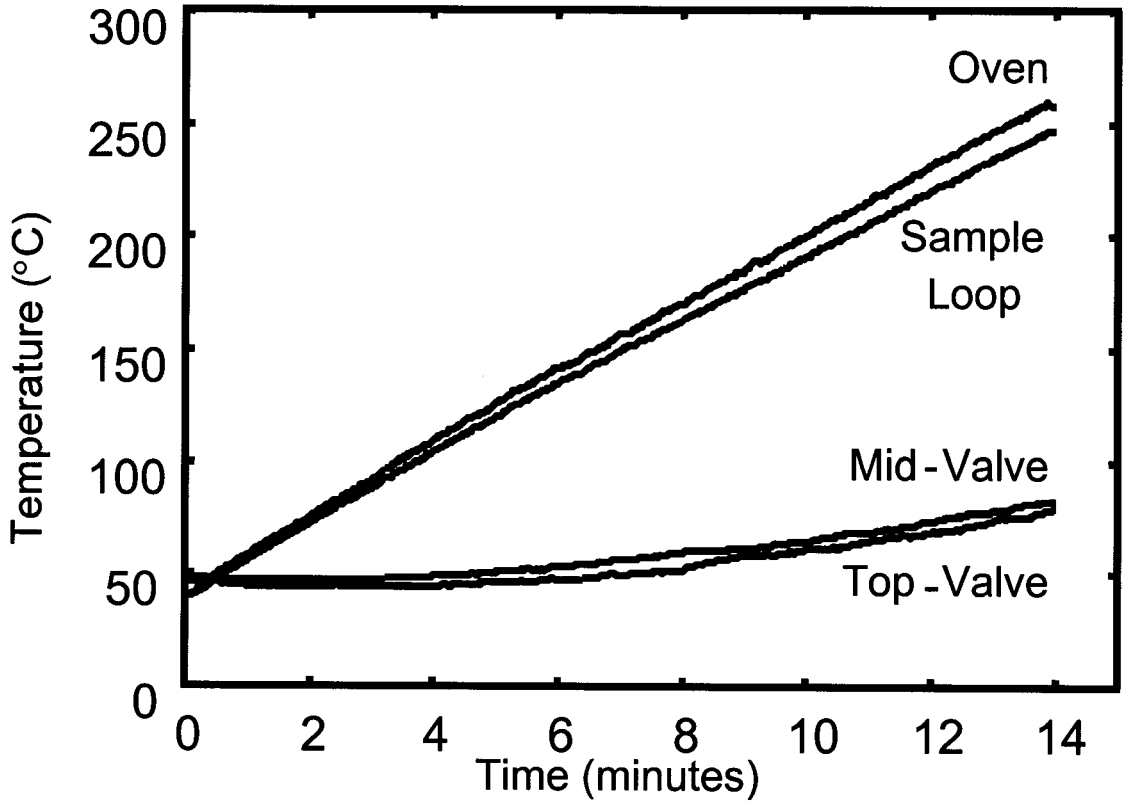


Figure 2.3A. Valve temperature profiles during ramp. The four type-K thermocouple sensors: Top-Valve, Mid-Valve, Sample Loop and Oven as indicated in Figure 2.2. The GC oven was programmed to ramp from 50 °C to 250 °C at 15 °C/min. Data was collected at 1 Hz.

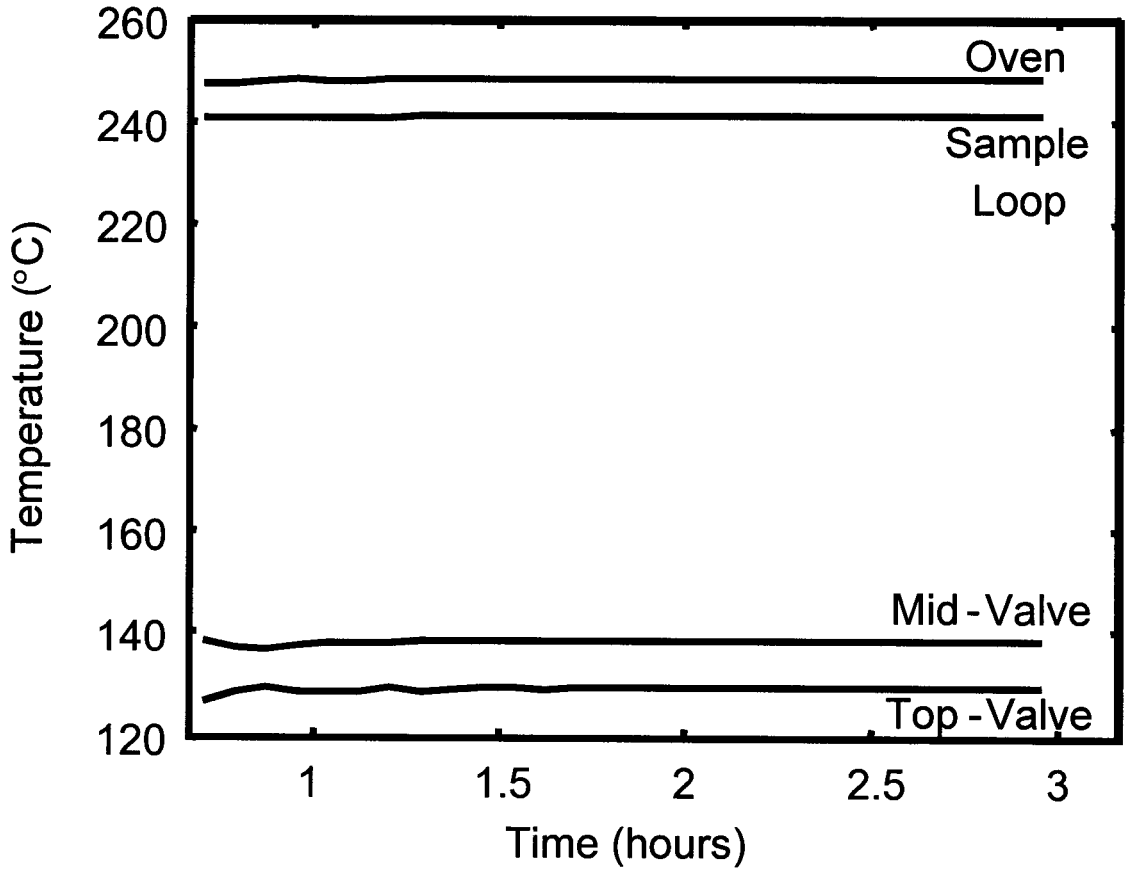


Figure 2.3B. Valve temperature profiles for high temperature hold. Data shown is for the four type-K thermocouple sensors when the GC oven is programmed to hold for three hours at 240 °C. Data was collected at 1 Hz and boxcar averaged every 5 minutes.

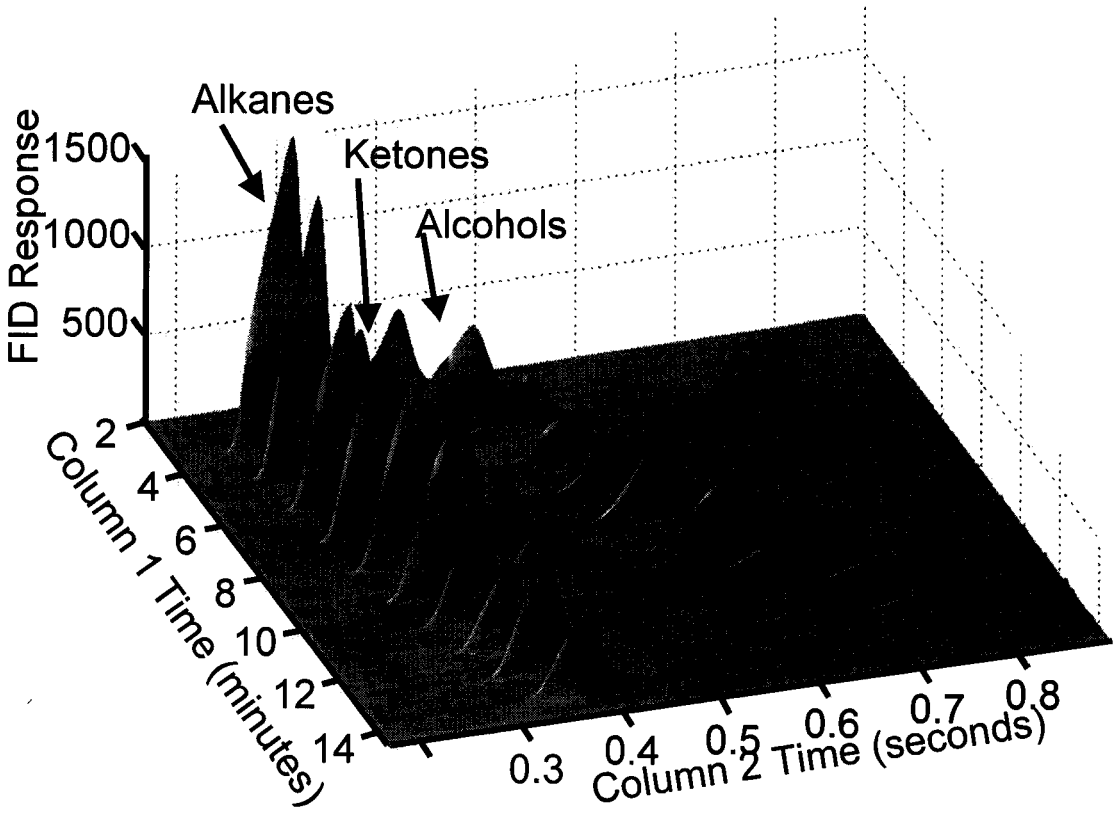


Figure 2.4A. Surface plot of the GC x GC separation of a 26-component test mixture. The mixture contained standards of homologous series of n-alkanes, 2-ketones and n-alcohols injected at equal volume percent listed in Table 2.1. Instrumental parameters: column 1 head pressure: 4 psi; column 2 head pressure: 20 psi; oven program: held at 40 °C for 1 minute and then programmed to ramp to 250 °C at 15 °C/min.

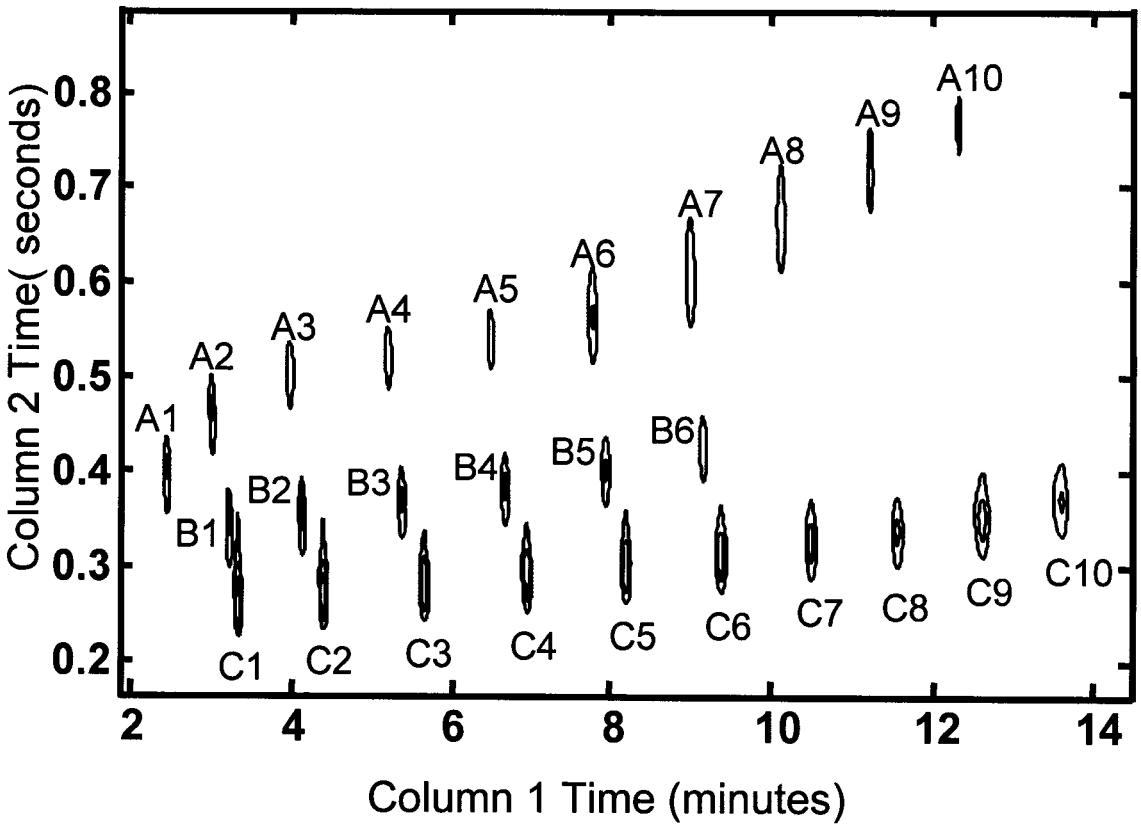


Figure 2.4B. Contour plot of the GC x GC separation of the 26-component mixture. **A1 - A10**: ethanol through n-undecanol; **B1 - B6**: 2-butanone through 2-nonanone; **C1 - C10**: n-hexane through n-pentadecane. Boiling points range from 65 °C (n-hexane) to 270 °C (n-pentadecane).

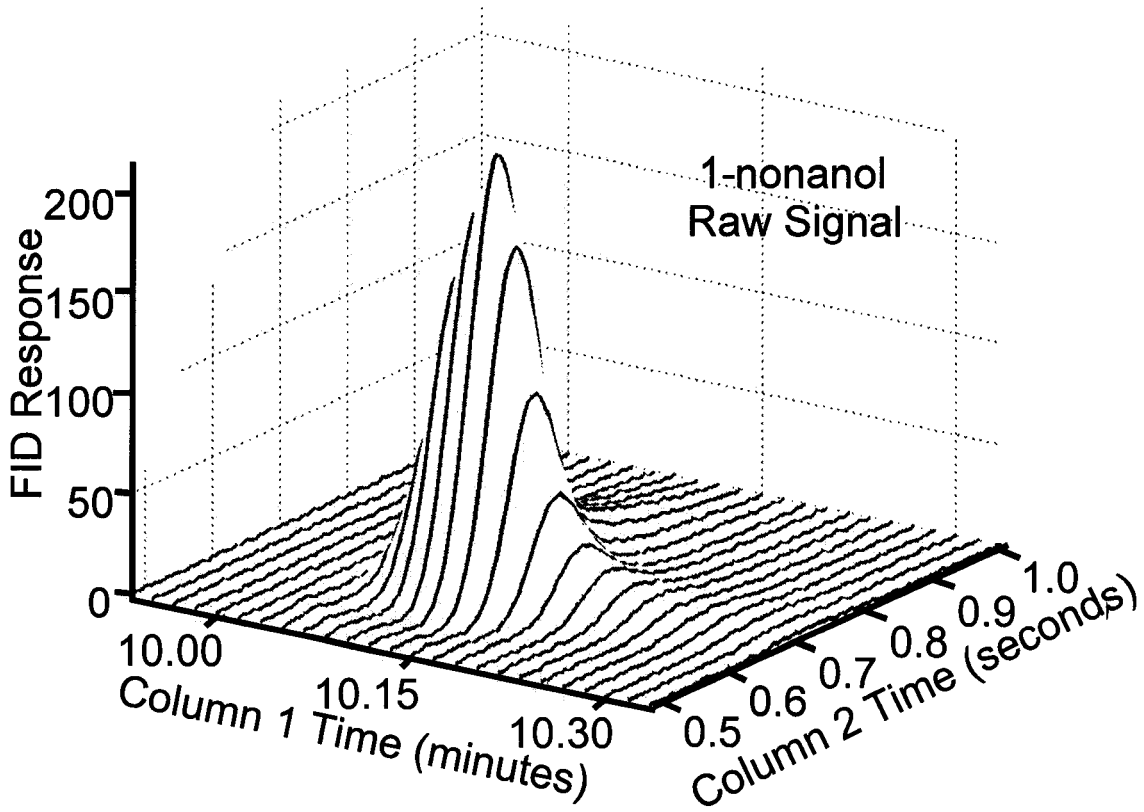


Figure 2.5A. 3D sub-matrix of 1-nonanol (**A8**) from the 26-component mixture.

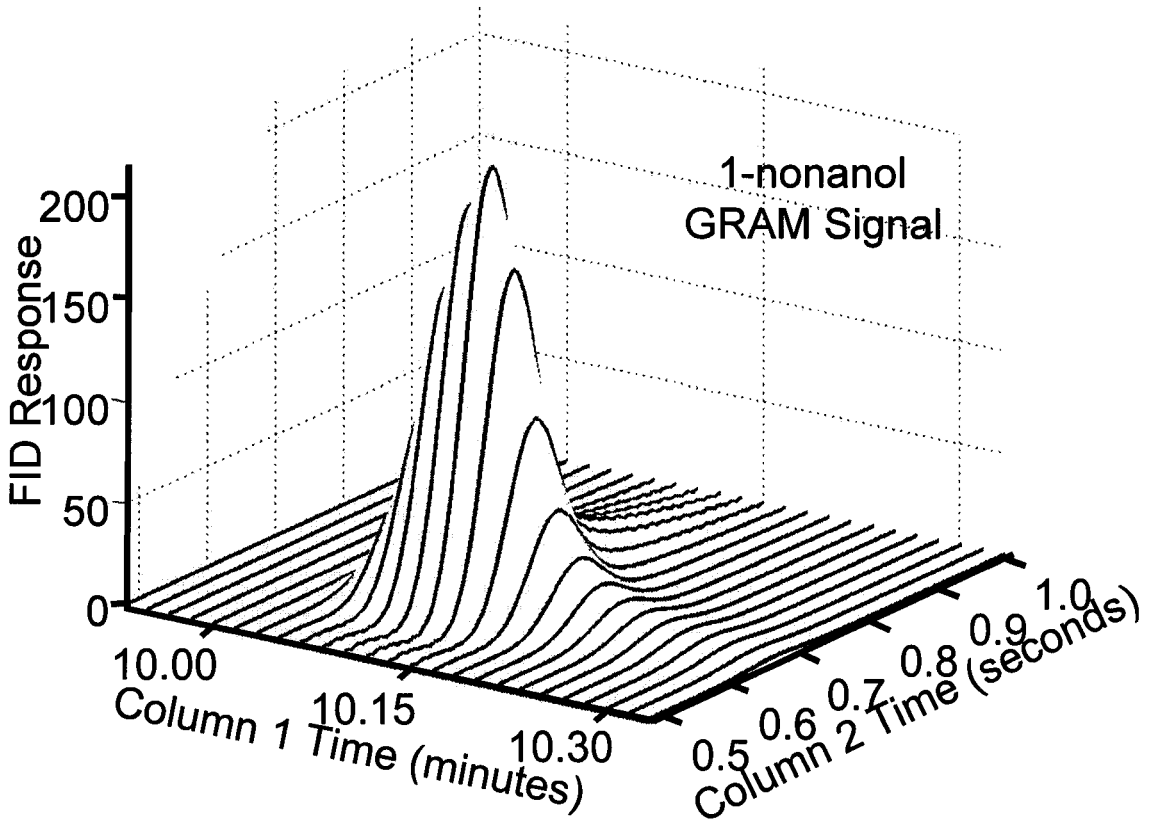


Figure 2.5B. 3D sub-matrix of 1-nonanol reconstructed from GRAM analysis. Representative data shown.

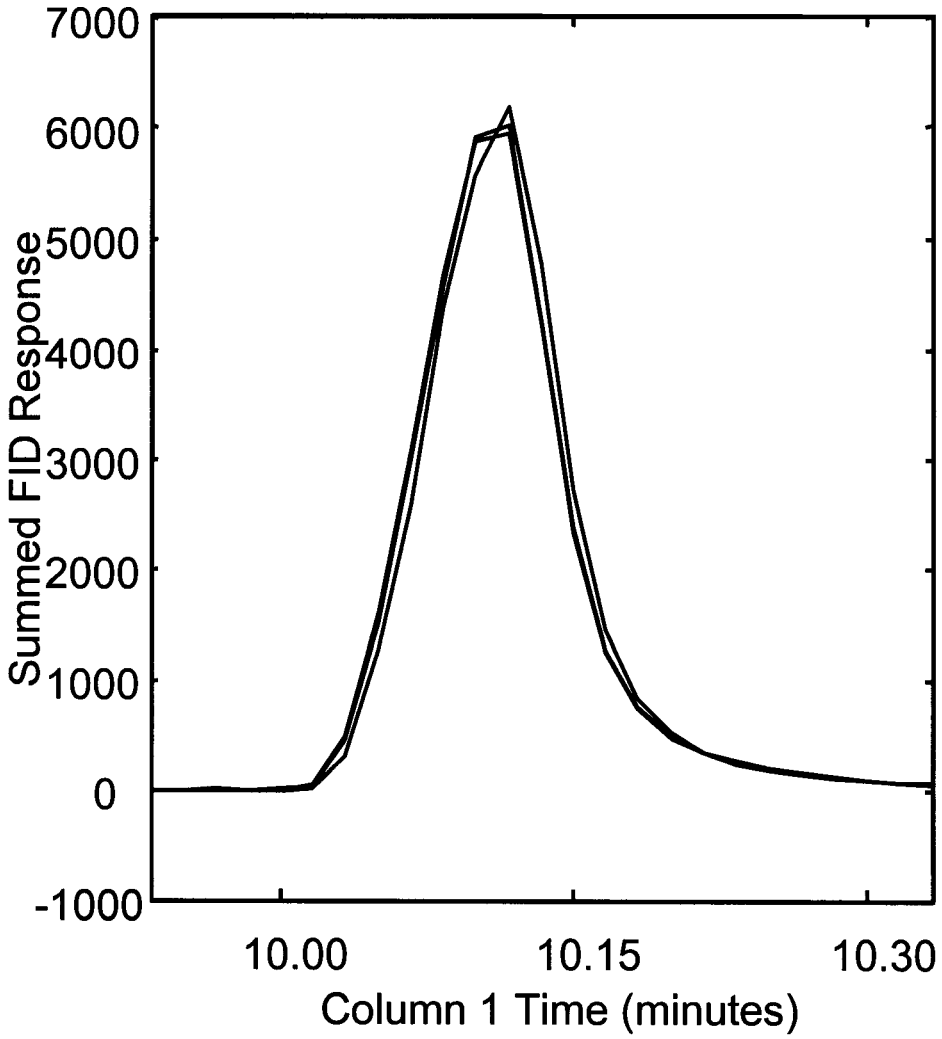


Figure 2.5C. Summed GRAM column 1 signal profiles for 1-nonanol. Results are from reconstructed sub-matrices of three replicate sample runs following application of GRAM.

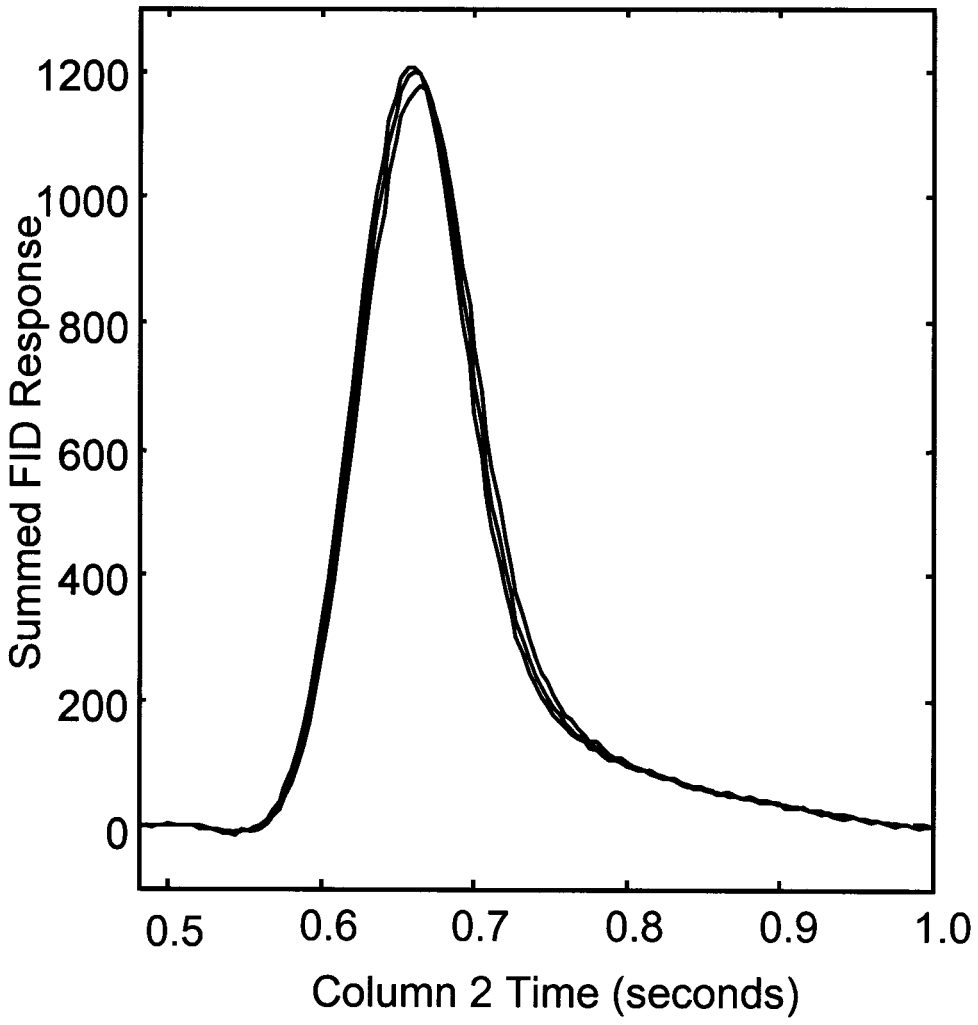


Figure 2.5D. Summed GRAM column 2 signal profiles for 1-nonanol. Results are from reconstructed sub-matrices of three replicate sample runs following application of GRAM.

CHAPTER 3. Valve-Based GC x GC with Time-of-Flight Mass Spectrometric Detection: Instrumentation and Figures-of-Merit*

3.1 Introduction

Comprehensive two-dimensional gas chromatography (GC x GC) is now a widely accepted powerful technique to analyze complex chemical mixtures and natural products such as petrochemicals, essential oils, fatty acids, and environmental pollutants [Beens et al. 2001b; Bertsch 2000; de Geus et al. 1998; de Geus et al. 2001; Dimandja et al. 2000; Fraga et al. 2000b; Frysinger et al. 1999; Frysinger and Gaines 2001; Harju and Haglund 2001; Johnson et al. 2002; Lee et al. 2001; Lewis et al. 2000; Marriott et al. 2000; Marriott and Shellie 2002; Prazen et al. 2001; Shellie et al. 2001a]. While the basic concept of comprehensive two-dimensional separations was laid out by Giddings [Giddings 1984; Giddings 1987], a recent summary of the state of comprehensive two-dimensional separations by Liu and Lee demonstrates how ubiquitous two-dimensional separations have become [Liu and Lee 2000]. The heart of a GC x GC instrument is a sample modulator that interfaces two chromatographic columns. The modulator injects aliquots of the first column effluent onto the second column at rapid intervals. In practice the sampling rate for a comprehensive two-dimensional separation should be such that all the components in the sample are analyzed by the second separation even if the entire injected volume is not analyzed [Bushey and Jorgenson 1990a; Bushey and Jorgenson 1990b]. In addition, peak shapes and resolution in the

---

* Large portions of this Chapter have been previously published [Sinha et al. 2003b].

first separation should also be preserved in the comprehensive two-dimensional separation, which can be achieved by sampling the first dimension a minimum of 3 to 4 times per peak [Murphy et al. 1998]. The first chromatographic column is the longer of the two columns, and often has a non-polar stationary phase resulting in a separation based primarily on boiling point. The second column is shorter, often with a polar stationary phase resulting in a separation based on relative polarity of the components. Thus, GC x GC separations utilize two columns providing complementary separations, which ultimately produce a separation space with a high degree of informational orthogonality [Liu et al. 1995; Slonecker et al. 1996; Venkatramani et al. 1996]. This leads to a greatly enhanced utilization of the available peak capacity, which is ideally equal to the multiplicative product of the peak capacity for each dimension [Giddings 1990]. GC x GC is therefore considered more powerful and, in many instances, less time consuming than traditional one-dimensional GC. With the exception of a few reports [Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; Shellie et al. 2001b; van Deursen et al. 2000], most GC x GC is done with flame ionization detection. In this study we report the development and evaluation of a valve-based GC x GC coupled to a time-of-flight mass spectrometer (TOFMS).

Two categories cover the different approaches for GC x GC modulation: thermal modulators and valve-based modulators. Most GC x GC instruments employ thermal modulation [Beens et al. 2001a; de Geus et al. 2000; Kinghorn and Marriott 2000; Korytar et al. 2002,; Ledford and Billesbach 2000; Marriott and Kinghorn 1997; Phillips and Xu 1995; Phillips et al. 1999], but there are a growing number of valve-based GC x GC instruments [Bruckner et al. 1998; Johnson et al. 2002; Prazen et al. 2001; Seeley et al. 2000; Seeley et al. 2002; Sinha et al. 2003a]. One recent advance in valve-based

GC x GC instrumentation was adaptation of a new valve-placement that extends the working temperature range of valve-based instruments to ~250 °C, which is 75 °C above the manufacturer's specified temperature limit of the valve [Sinha et al. 2003a]. This advance has extended the applicability of valve-based GC x GC instruments to semi-volatile components. Utilization of this new high-temperature configuration is reported in this chapter, in conjunction with TOFMS detection, for the analysis of selected components in a complex environmental sample. Valve-based GC x GC instruments offer a simple, rugged and less expensive alternative to thermally modulated units and have the potential to be portable. A recent instrumentation advance also reported in this chapter includes a stand-alone pulse generator to power and control the valve that eliminates the need for a computer and LabVIEW software to drive the valve-based modulator. This development reduces the cost of the valve-based GC x GC and provides an opportunity to simplify the construction of a field-portable GC x GC.

This work was the first report of a GC x GC-TOFMS using a valve-based modulator. By coupling mass spectrometry to GC x GC, identification of components is possible using mass spectral database searches or a handful of highly selective ions for specific components. Conclusive identification of analytes without the use of standards is often not possible with commonly used universal detectors such as flame ionization detection (FID). The most common types of mass spectrometers for GC detection have been quadrupole instruments. The scanning speed of most quadrupole mass selective detectors (MSD) (~0.5 second for *m/z* 40 to 400) [Shellie et al. 2002] is not fast enough to use as a detector for GC x GC instruments in which the second column peak widths are routinely around 100 milliseconds. GC x GC with MSD can be used for qualitative analysis using limited mass range scanning, but the

precision of the quantitative analysis is compromised at the slow scanning rate. Recently available TOFMS analyzers that are able to store 500 full spectra per second are perfectly suited for GC x GC detection and have been demonstrated with thermal modulation instruments [Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; Shellie et al. 2001b; van Deursen et al. 2000]. The pressure and flow conditions for thermal modulation and valve-based GC x GC instruments are substantially different providing some unique coupling challenges for valve-based instruments to TOFMS detectors. Therefore, the development and evaluation of a valve-based GC x GC-TOFMS instrument is reported herein. A complex environmental sample was analyzed to investigate the separation and performance characteristics of this novel system.

3.2 Experimental

An Agilent 6890 Gas Chromatograph equipped with an Agilent 7683 Auto-injector (Agilent Technologies, Palo Alto, CA, USA) was modified to a valve-based GC x GC by mounting the wetted portions of the high-speed six-port micro diaphragm valve (VICI, Valco Instruments Co. Inc. Houston, TX, USA) inside the oven and the remaining portions outside the oven [Sinha et al. 2003a]. The second column was connected to a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI, USA) via the heated transfer line (Figure 3.1). A GC x GC-TOFMS analysis was performed on a complex environmental sample (#1). Environmental sample #1, an acetone extract from a contaminated landfill site of a proprietary nature, originally contained a mixture of fuel components and natural products. Three organo-phosphorous pesticides: triethyl phosphorothioate (TEPT) (AccuStandard, Inc., New Haven, CT, USA), dimethyl methyl phosphonate (DMMP) (Sigma-Aldrich Corp., St. Louis, MO, USA), and dimethyl phosphite (DMP) (Sigma-Aldrich Corp.) were

spiked into environmental sample #1. These compounds can sometimes be found in similar environmental samples; however, they were not detected prior to addition in this particular sample.

The first column of the GC x GC-TOFMS was a 60-m 250-μm i.d. capillary column with a 0.5-μm 5% diphenyl/95% dimethyl polysiloxane film (DB-5; J&W Scientific/Agilent Technologies). The second column was a 3-m 180-μm i.d capillary column with a 0.05-μm 90% biscyanopropyl/10% phenylcyanopropyl film (RTX-2330; Restek Corp., Bellfonte, PA, USA). Helium was used as the carrier gas. Column 1 was operated with a constant flow of 1.0 ml/min. Column 2 was operated with a constant pressure of 20 psi. The injector set point was 275 °C and 2.0 μl injections of the sample were splitless for 0.5 minutes. The oven was held at 40 °C for 0.5 minutes, ramped to 80 °C at 20 °C/min, ramped to 210 °C at 5 °C/min, then ramped to 230 °C at 20 °C/min and held for 4 minutes. The valve coupling column 1 to column 2 was equipped with a 5 μl sample loop and actuated every 2.5 seconds with a 60-ms injection pulse width. Under these experimental conditions, it is estimated that 10% of the sample is transferred from the first column to the second column. Essentially, this is equivalent to a 1:10 split at the injector onto column 1. A stand-alone pulse generator that was designed and built in-house was used to control the valve actuation (Valve Controller in Figure 3.1). This new hardware allows the user to set the pulse width, the period of the actuations, and the total duration of the analysis. This new development in the valve-based GC x GC instrument replaces the LabVIEW program and counter/timer board thus reducing the cost and enabling field portability. The mass spectrometer had a transfer line temperature of 250 °C and an ion source temperature of 200 °C. The filament bias voltage was -70 V and the detector voltage was -2000 V. All other MS parameters were set from the

results of an automatic optimization sequence controlled by the LECO software using perfluorotributylamine (PFTBA) as the standard. Data were collected from *m/z* 40 to 300 at a nominal rate of 5 kHz and averaged to 50 full spectra/second by the LECO software, resulting in more than 4 mass spectra per second-dimension peak (column 2). Data were then exported as a comma separated value (.csv) file and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing.

3.3 Results and Discussion

Figure 3.2A depicts the two-dimensional total ion current (TIC) chromatogram for environmental sample #1 analyzed for the purpose of investigating instrumental performance. The complexity of the sample was such that a 30-minute first column separation time was required. A column 2 separation time of 2.5 seconds was used. The column 2 flow rate is constrained by the 10-ml/min maximum flow allowed for the TOFMS. With these parameters, at least 4 injections onto column 2 during the elution of a peak from column 1 were achieved. Since the TIC is a sum of all of the mass channels at each point in the two-dimensional separation space, only the largest peaks are visible in this image. Note that high-concentration compounds eluting at about 11-12 minutes overload column 2, resulting in the broad peaks with apparent breakthrough. A peak location plot for the TIC is shown in Figure 3.2B. The TIC plot was generated by an algorithm written in-house that locates and marks peak maxima. A representative GC x GC chromatogram and a peak locator plot of *m/z* 99 for this sample are shown in Figure 3.2C and D in order to further illustrate the sample complexity and near-full use of available peak capacity. It can be seen that the individual *m/z* 99 mass channel and the TIC chromatograms exhibit different patterns of peaks. It can also be seen from *m/z* 99 that a there are peaks filling a large

percentage of the separation space. The peak capacity of this separation space was estimated to be ~2000 using a resolution of 1. The *m/z* 99 chromatogram was chosen because it contains a pattern of peaks that are not seen in the TIC chromatogram, further illustrating the use of nearly all the available peak capacity. Other mass channels exhibit different peak patterns emphasizing the high selectivity of GC x GC-TOFMS and extreme complexity of this sample, but were not shown for brevity.

Figure 3.3 is a sub-region of the TIC chromatogram containing approximately 20 peaks. This figure illustrates the separation efficiency of the valve-based GC x GC coupled with TOFMS in greater detail than the previous figures. As can seen in Figure 3.3, the nominal peak widths at the base in the column 2 dimension are from 100 to 300 milliseconds, thus indicating an efficient column 2 separation, similar to what is achieved with thermal modulation-based GC x GC systems.

The injected mass and injected concentration limit of detection (LOD) for three organo-phosphorous compounds (TEPT, DMMP, and DMP), were determined using the standard addition method by spiking known amounts of the pure components into environmental sample #1 (Table 3.1). The locations of the target compounds are not directly identified in any of the GC x GC plots in Figure 3.2, however their retention times are listed in Table 3.1. Three replicate data sets were collected in order to gauge the reproducibility of the instrument. The LOD is based on the most abundant and/or the most selective mass channels for each analyte (Table 3.1). The three-dimensional signal region for *m/z* 198 containing TEPT is shown in Figure 3.4A. The signal intensity was determined as the height of the largest column 2 peak above the mean baseline (Figure 3.4B). The LOD for each analyte was defined as 3σ where σ was determined from the average standard

deviation of four column 2 chromatograms (i.e., sections of baseline, absent of peaks) adjacent to the peak consisting only of baseline noise. The injected mass and injected concentration LODs were from 6 to 38 pg, and 3 to 17 ng/ml, respectively. This LOD range is comparable to results recently reported using thermal modulation GC x GC-TOFMS [Dallüge et al. 2002c] for trace analysis.

An experimentally obtained mass spectrum for TEPT at ~ 25 times the LOD and the resultant identity search NIST library spectrum show a significant degree of similarity even at such a low concentration (Figure 3.4C and D). There are multiple factors that one can use to evaluate the quality of a spectral match in the NIST MS Search program. The match factor evaluates how closely the target spectrum and the database spectrum correlate. It is calculated by taking the inner product of the two spectra, with lower *m/z* peaks having less weight than higher *m/z* peaks. Values are reported on a scale from 1-1000 where a perfect match is 1000. The reverse match factor ignores impurity peaks in the experimental spectrum, that is, peaks that are not present in the library spectrum. This is also reported on a scale of 1-1000. The probability of the unknown spectrum arising from the same compound that generated the library spectrum is listed on a scale from 1-100. For the experimental spectrum of TEPT (Figure 3.4C), a library search resulted in a match value of 735, a reverse match of 793, and a probability of 78.7. Extracted spectra at higher concentrations of TEPT result in higher match factors, but considering the low concentration the spectral match quality is very high. Along with the speed that makes TOFMS amenable to GC x GC detection, the added selectivity of full spectral detection lends more credence to analyte identification. Mass spectral analysis of the other two pesticides was also successful with the results not shown for brevity. If an analyst

requires higher match values, it may be prudent to design your own library with analyte standards. This would minimize spectral variation due to differences in the ionization source for the NIST system relative to the TOFMS in use. Indeed, ionization source difference is the cause for the match values with the TEPT study falling below the ideal.

3.4 Conclusions

It was demonstrated that valve-based GC x GC-TOFMS is a selective and sensitive instrument for the analysis of complex samples. Detection limits for organo-phosphorous compounds in a complex sample were determined to be between 6 and 38 pg and typical column 2 peak widths ranged from 100 to 300 milliseconds. Furthermore, the GC x GC-TOFMS instrument provides the unambiguous identification of compounds based on their full mass spectra. In addition to the attributes just listed, valve-based GC x GC-TOFMS provides the opportunity to use unique chemometric signal deconvolution techniques. Previous work has demonstrated GC x GC peak deconvolution using techniques like generalized rank annihilation method (GRAM) [Bruckner et al. 1998; Fraga et al. 2001a; Fraga et al. 2001b; Johnson et al. 2002; Prazen et al. 2001]. GRAM and other chemometric deconvolution techniques utilize the bilinear structure of GC x GC data to separate unresolved chromatographic peaks. GRAM requires the comparison of a sample data set and a standard data set for deconvolution and quantitation. The unique structure of GC x GC-TOFMS data, which is known as third-order data (i.e., a trilinear data structure), allows for signal deconvolution without the comparison of data sets [Booksh and Kowalski 1994]. This eases the retention time precision requirements, and thus simplifies signal deconvolution.

Table 3.1. Analytical figures-of-merit for three organo-phosphorous pesticides. Triethyl phosphorothioate (TEPT), dimethyl methyl phosphonate (DMMP), and dimethyl phosphite (DMP) determined from three replicate samples analyzed with standard addition. Injected concentrations of the analytes of interest ranged from 25 to 50 times the LOD. The limit of detection was determined from the peak height of the maximum column 2 chromatogram and 3 times the average standard deviation of 4 adjacent rows of noise. The *m/z* used for quantitation was the most abundant and/or selective ion for each analyte. The limit of detection is listed in concentration injected and in absolute mass injected for a 2 μl injection. Peak widths were measured at peak base. For each entry the standard deviation is also indicated.

	m/z	Inj. Conc. (ng/ml)	Col. 1 R_t (min)	Col. 2 R_t (s)	Col. 1 width (s)	Col. 2 width (ms)	LOD (3σ) (ng/ml)	LOD (3σ) (pg)
DMP	47	878	10.51 ± 0.02	1.037 ± 0.005	7.4 ± 0.6	218 ± 10	17 ± 3	38 ± 6
DMMP	94	273	12.25 ± 0	1.04 ± 0	7.4 ± 0.7	293 ± 50	7 ± 1	14 ± 2
TEPT	198	70	20.38 ± 0	0.403 ± 0.005	8.8 ± 0.5	156 ± 19	3.2 ± 0.2	6.4 ± 0.4

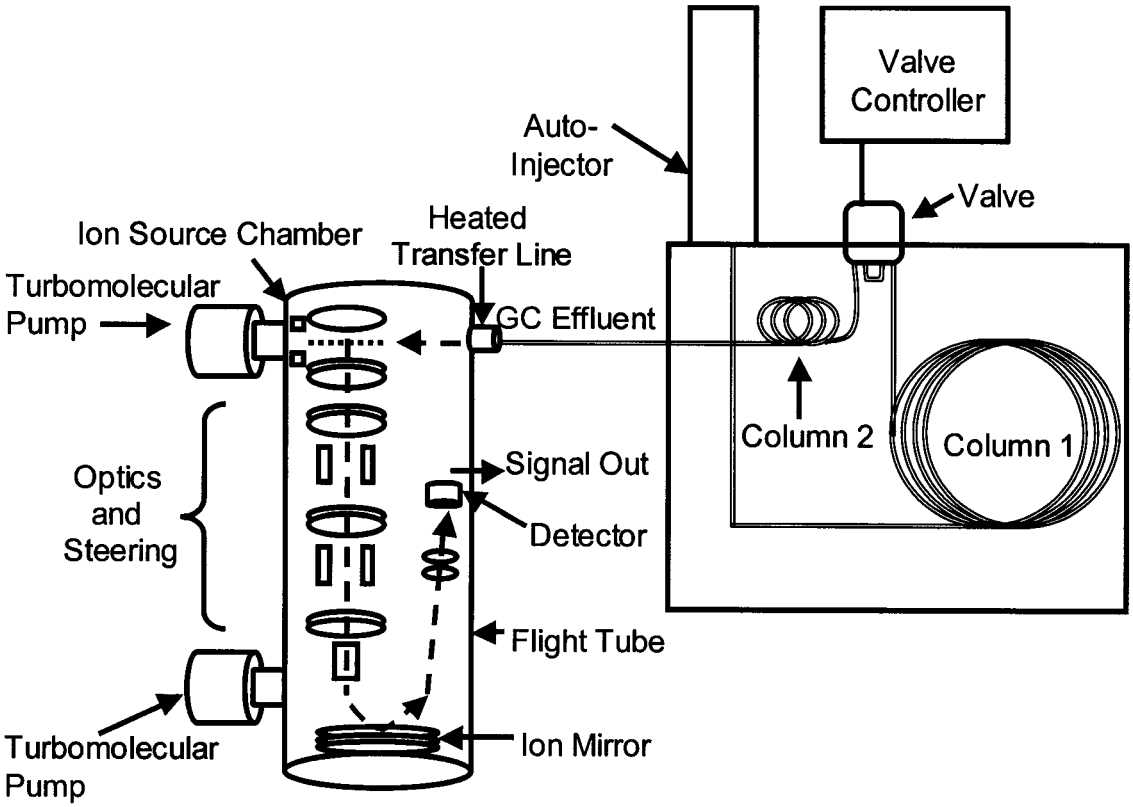


Figure 3.1. Valve-based GC x GC-TOFMS instrument schematic. Both columns are housed inside the same oven. A stand-alone pulse generator (valve controller) is utilized in the system to simplify the instrumentation and to promote portability. Modulation occurs via a sample loop on a 6-port high-speed mini-diaphragm valve.

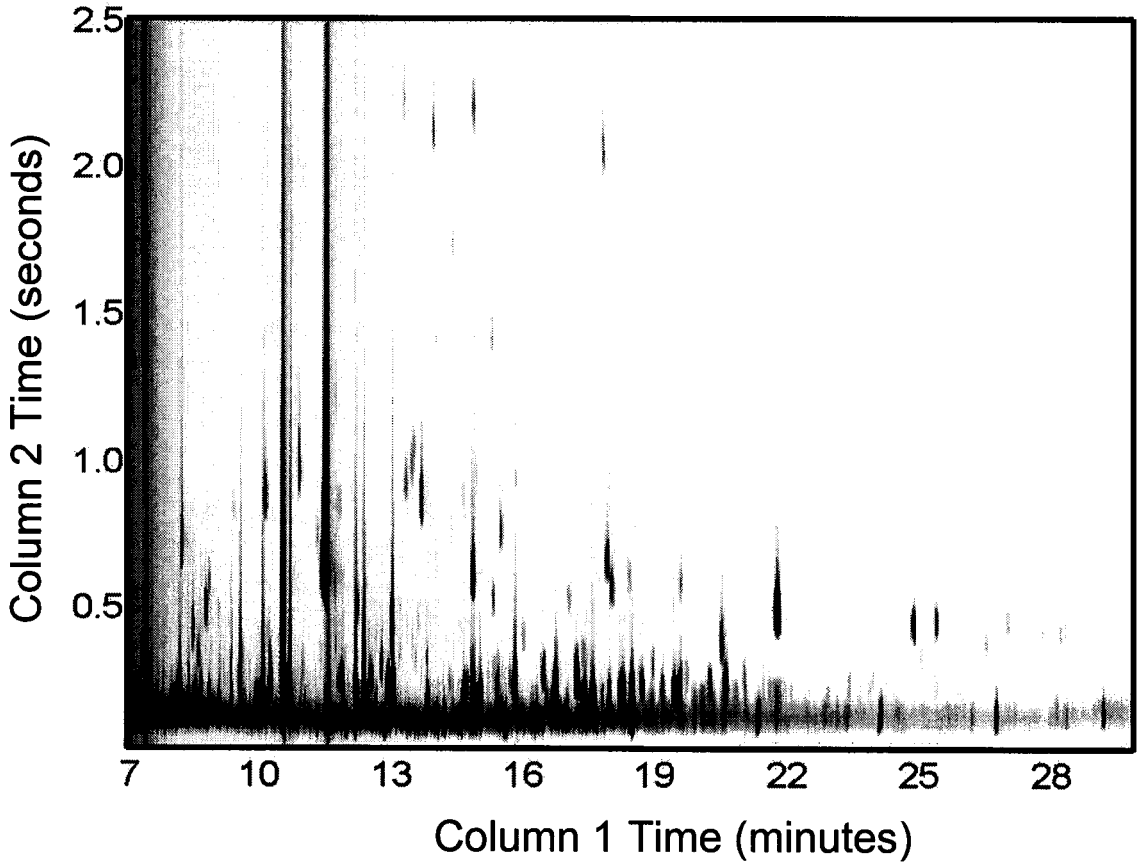


Figure 3.2A. TIC chromatogram of complex environmental sample #1.

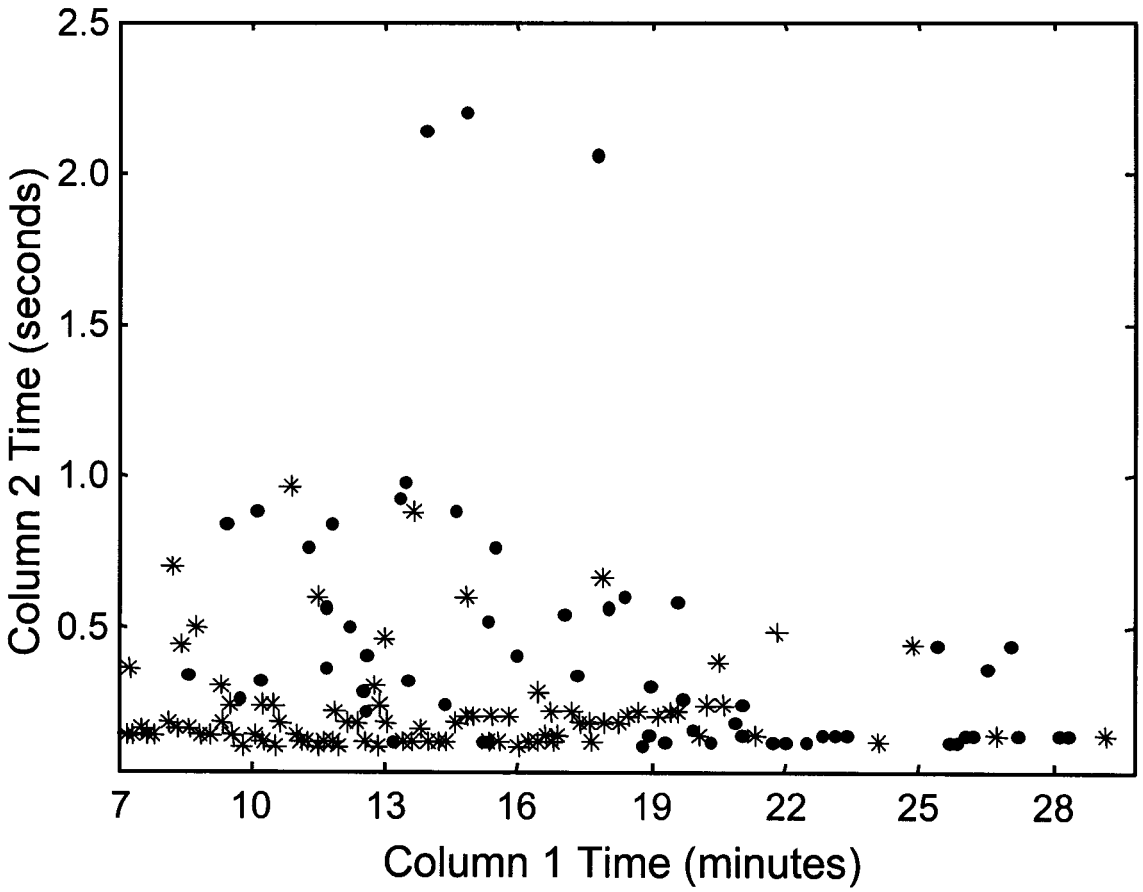


Figure 3.2B. Peak locations in the TIC chromatogram. This image was generated with a program that identifies local peak maxima: * indicates peaks of heights 40%-100% of the maximum peak height (i.e., largest peak in the chromatogram), and • indicates peaks with heights 5σ to 40% of the maximum peak height.

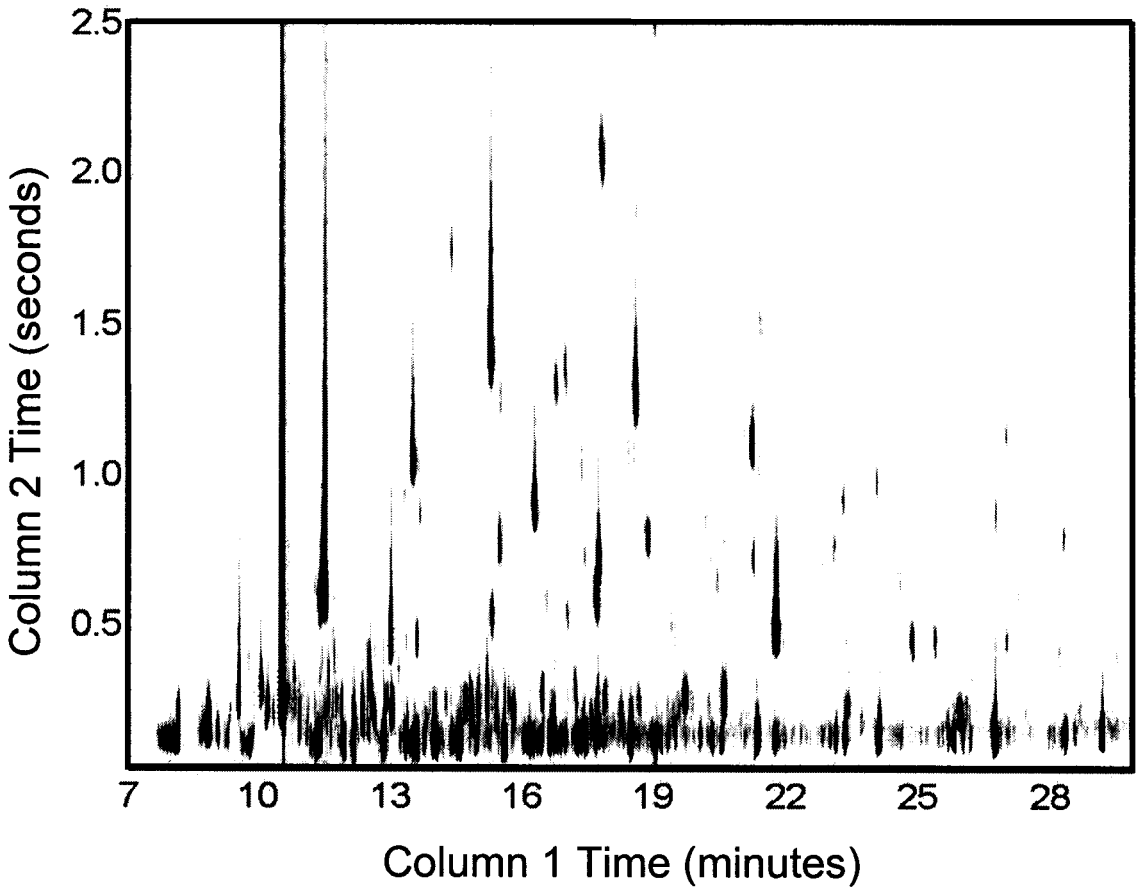


Figure 3.2C. Chromatogram of *m/z* 99. This demonstrates the selectivity of GC x GC-TOFMS in that this chromatogram and the TIC have very different peak distributions.

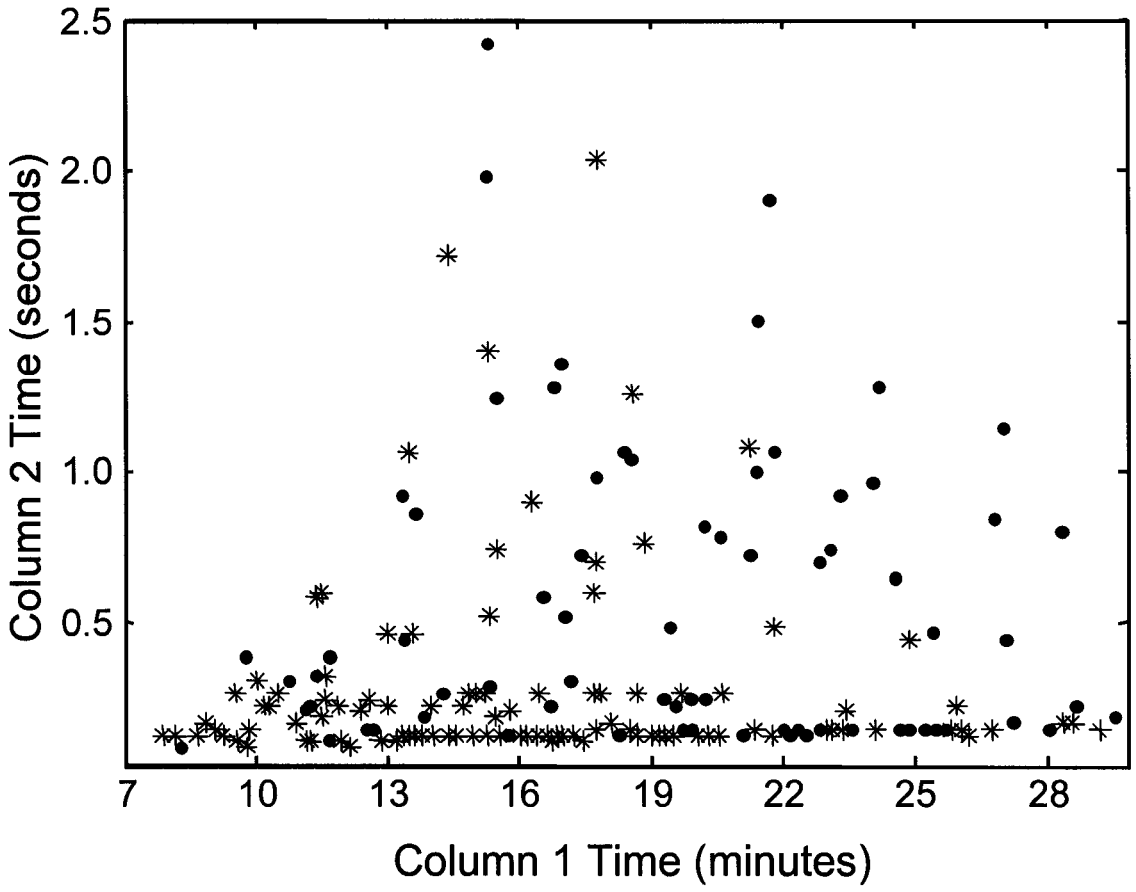


Figure 3.2D. Peak locations in the chromatogram of *m/z* 99. This image was generated with a program that identifies local peak maxima: * indicates peaks of heights 40%-100% of the maximum peak height, and • indicates peaks with heights 5σ to 40% of the maximum peak height.

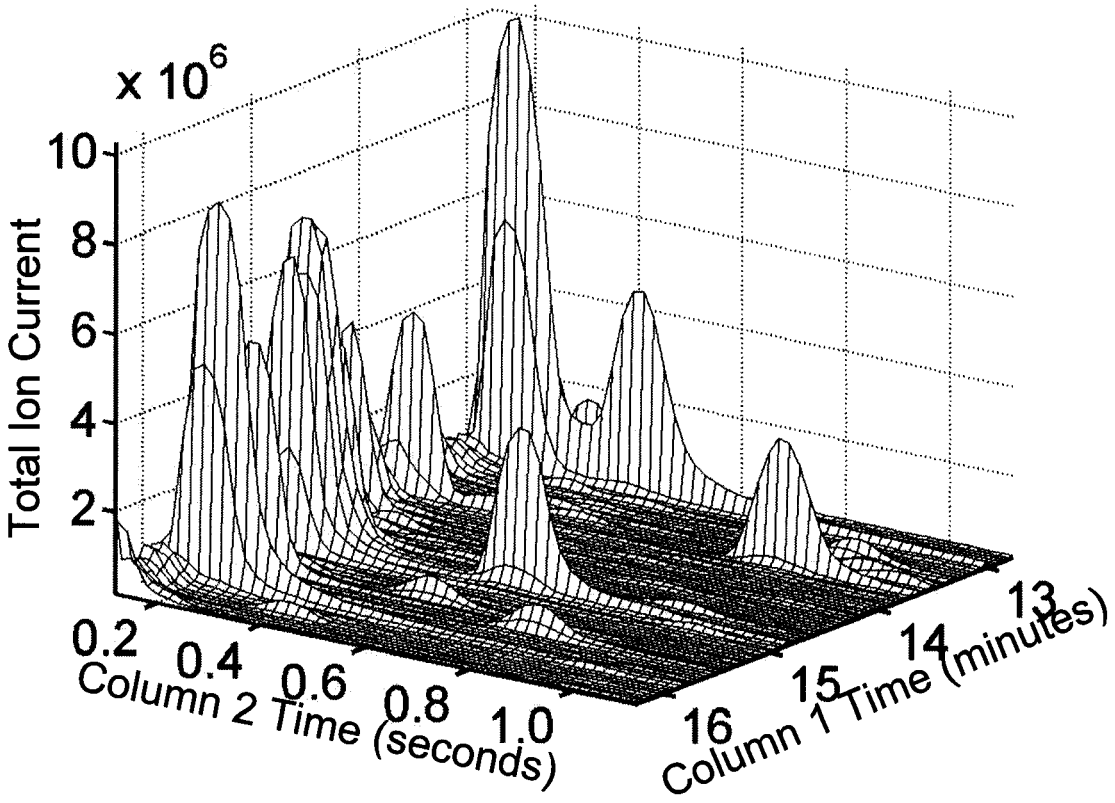


Figure 3.3. Subsection of environmental sample #1 GC x GC chromatogram. The separation efficiency and resolution are clearly illustrated. Note that the axis is rotated to make more peaks visible.

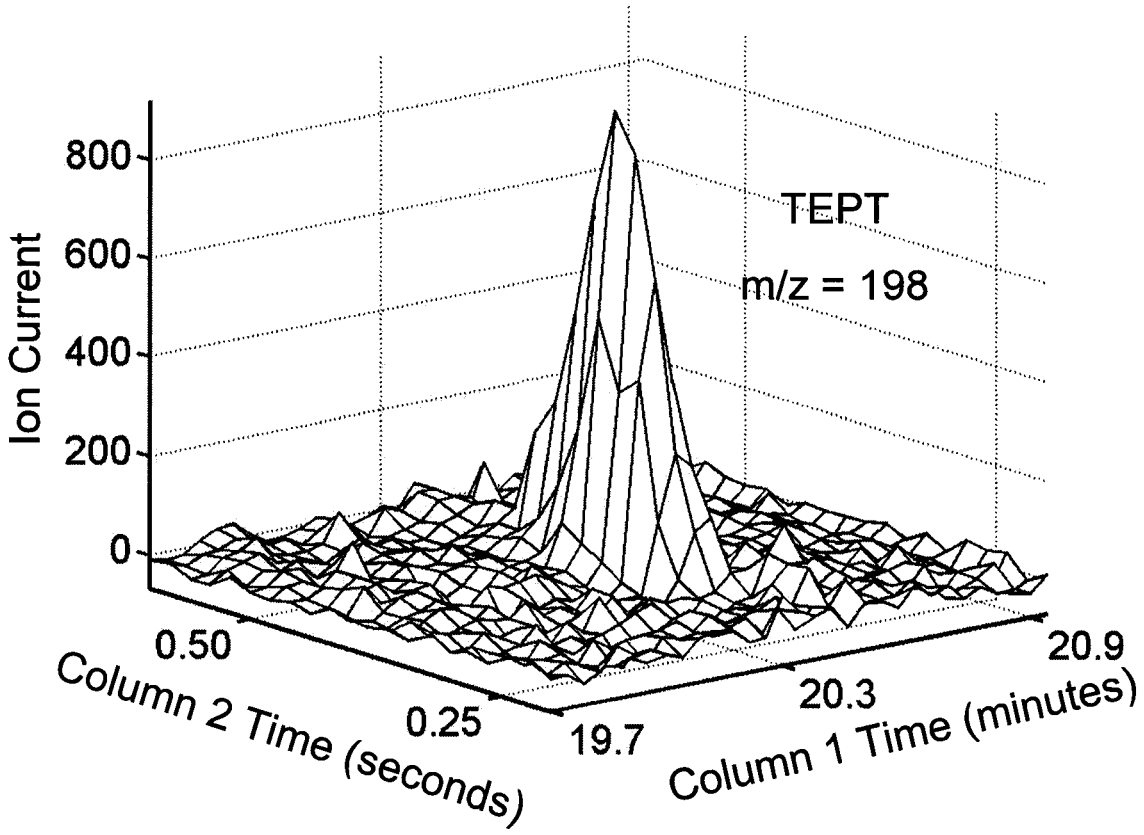


Figure 3.4A. 3D region of chromatogram of *m/z* 198 of TEPT peak. Injected concentration was 70 ng/ml.

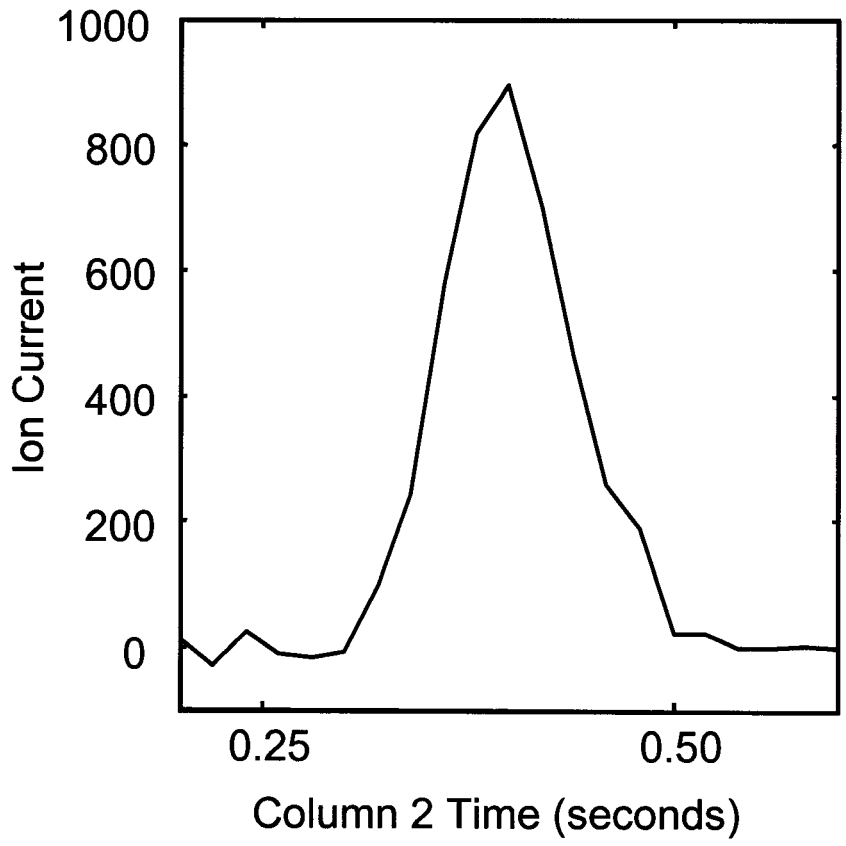


Figure 3.4B. Highest column 2 chromatogram of *m/z* 198 for the TEPT peak. LOD was calculated at 3σ using an average of the standard deviations of four adjacent column 2 chromatograms consisting entirely of baseline noise. The LOD was calculated to be 3 ng/ml.

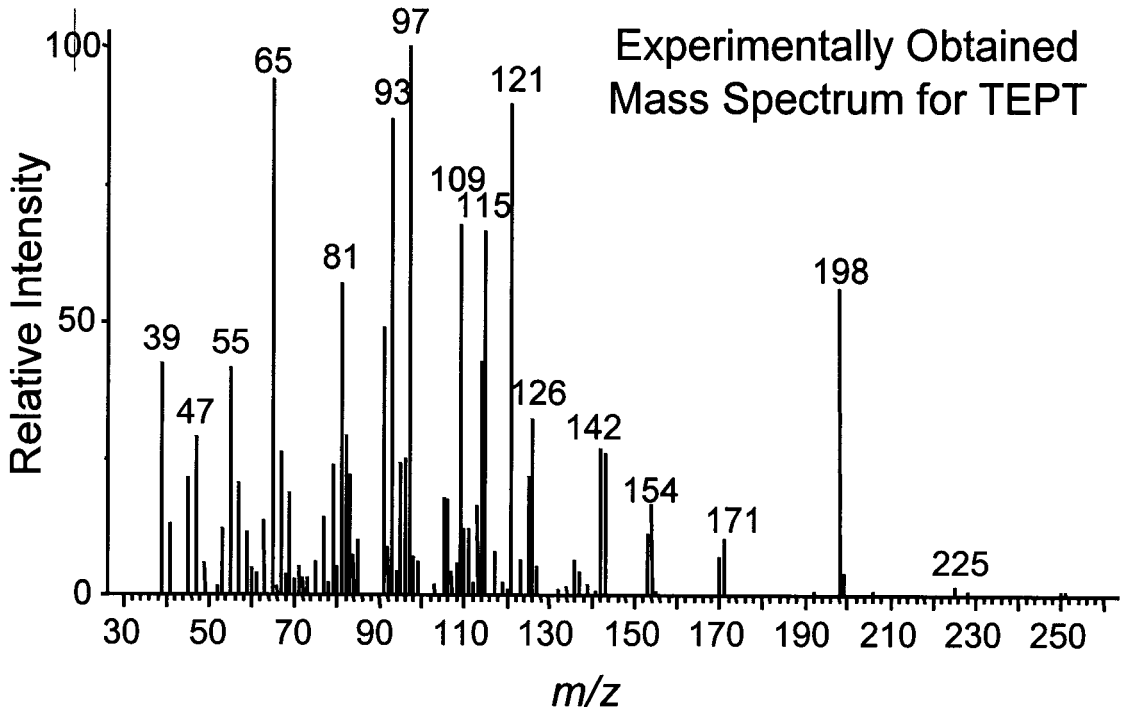


Figure 3.4C. Experimentally obtained spectrum of TEPT at 70 ng/ml.

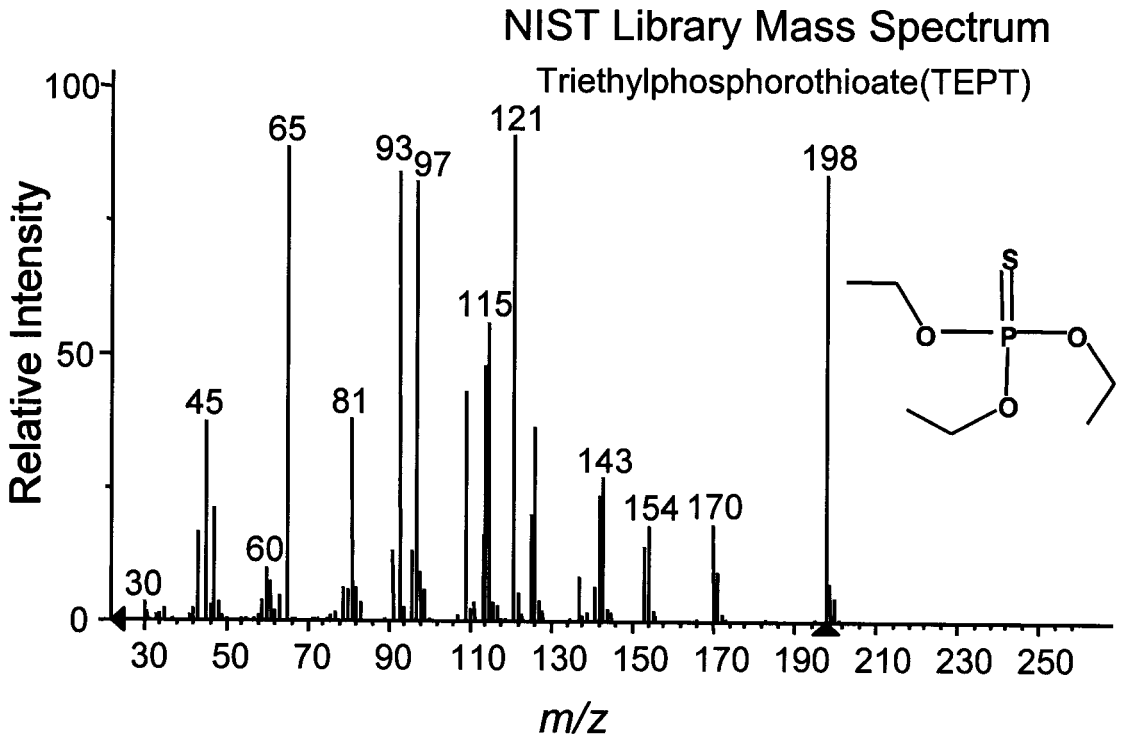


Figure 3.4D. NIST Library matched spectrum for TEPT.

CHAPTER 4. Trilinear Chemometric Analysis of GC x GC - Time-of-Flight Mass Spectrometry Data*

4.1 Introduction

Recently, the combination of comprehensive two-dimensional gas chromatography (GC x GC) with time-of-flight mass spectrometry (TOFMS) has become an active area of research [Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; Shellie et al. 2001b; Sinha et al. 2003b; van Deursen et al. 2000]. Not only is GC x GC-TOFMS an excellent instrument for the identification of components in complex mixtures, but it is capable of generating trilinear data, thus broadening the opportunity to utilize state-of-the-art chemometric signal deconvolution techniques. It is essential to utilize this resulting trilinear data structure in order to realize the full power of the GC x GC-TOFMS technology and to optimize the extraction of useful information from each complex sample or from complex regions of otherwise simpler samples.

A trilinear data structure is a relatively high level structure [Booksh et al. 1994]. Like non-negative data structure, unimodal data structure and linear data structure, trilinear data structure can be a useful tool for the interpretation of chromatographic data. The trilinear data structure as provided by the GC x GC-TOFMS instrument is described in the Theory section. Qualitative and quantitative chemometric techniques continue to become more important tools for analytical chemists as analytical techniques are hyphenated and instruments continue to collect data faster than their predecessors. GC x GC-TOFMS is a perfect example of a state-of-the-art

---

* Large portions of this Chapter have been previously published [Sinha et al. 2004a].

hyphenated instrument that creates a seemingly unmanageable amount of data, yet has the potential to more fully describe the contents of complex chemical mixtures.

Current peak deconvolution methods typically used for GC x GC-TOFMS data analysis are those that are already used for GC-MS. These methods require a purely selective mass channel and peak width/shape estimates. Furthermore, the current GC-MS peak deconvolution methods that are applied to GC x GC-TOFMS data resort to reducing the data into a series of 2nd dimension GC-MS data, doing the analysis of each "2nd separation dimension slice" independently, and then recombining these results along the 1st separation dimension for the final assessment. However, with the added third dimension of selectivity and separation as in GC x GC-TOFMS, only partial selectivity is needed in both separation dimensions and the mass spectral dimension for peak deconvolution using a single data set, and the information along all three dimensions is analyzed simultaneously. Herein, we report that the trilinear data structure naturally provided by GC x GC-TOFMS combined with the appropriate chemometric techniques allows for deconvoluted chromatographic profiles and mass spectra to be produced without specifying selective mass signals and without resorting to requiring peak shape and width estimates. This work gives promise to future studies that will investigate the power of combining GC x GC-TOFMS with pattern recognition and multivariate quantitative classification methods that take advantage of the trilinear structure of the GC x GC-TOFMS data [Johnson and Synovec 2002; Prazen et al. 2001].

Chemometric calibration techniques such as the generalized rank annihilation method (GRAM), trilinear decomposition (TLD) and parallel factor analysis (PARAFAC) were developed to find relationships in trilinear

data [Booksh et al. 1994]. One method of generating trilinear data is to acquire a set of bilinear data, e.g., a set of multiple GC x GC analyses of a compound of interest at more than two concentrations. GRAM is used for analyzing two bilinear data samples at a time, while TLD was developed as an extension of GRAM, with TLD being capable of analyzing multiple bilinear data samples at once [Booksh et al. 1994; Gui et al. 1995; Sánchez and Kowalski 1986; Sánchez and Kowalski 1988; Sánchez and Kowalski 1990; Wilson et al. 1989]. All of these calibration methods require that the data structure be trilinear, which implies that "the response in [all] domains of the instrument arising from a species should be unique, consistent, and independent of the presence of other species" [Lin et al. 1994]. When both chromatographic columns in GC x GC are subjected to temperature programming, the bilinear data integrity can come into question and must be substantiated [Johnson et al. 2002].

With analytical instruments like GC x GC with flame ionization detection, calibration methods such as GRAM are able to deconvolute unknowns using two data sets (standard and sample) where the analytes of interest vary in concentration between the two data sets [Bruckner et al. 1998; Fraga et al. 2000a; Fraga et al. 2000b; Fraga et al. 2001a; Prazen et al. 1999a; Sánchez and Kowalski 1988; Synovec et al. 2003b]. Using trilinear data, such as GC x GC-TOFMS data, it is possible to deconvolute individual components from a group of partially overlapped components using a data set from only one sample. This ability to deconvolute a single data set with partially resolved signals into the fully resolved signals is known as the third-order advantage [Booksh and Kowalski 1994]. In the case of chromatographic data, third-order data is also advantageous because it relaxes the requirements for sample-to-sample retention time precision thus essentially eliminating the

need for retention time alignment prior to analyte deconvolution, identification, and quantification involving standard addition analysis.

In order to facilitate development of trilinear chemometric methods, an environmental sample containing fuel components, pesticides and natural products was separated with GC x GC-TOFMS. With the goal of demonstrating the chemometric techniques, not complete sample characterization, one representative region of overlapping peaks in the real, complex sample was analyzed with TLD and PARAFAC. It will be demonstrated that TLD and PARAFAC are able to deconvolute from initially partially resolved data, the pure component profiles in both chromatographic dimensions and the pure mass spectrum of each component. In addition, a procedure to identify and quantify components of interest in a complex mixture via a standard addition method is outlined. This analysis procedure is successfully demonstrated for an analyte of interest in the environmental fuel sample without requiring retention time alignment between sample and standard addition cases.

4.2 Theory

4.2.1 Trilinear Data

Mathematically, the trilinear PARAFAC model is described as

$$\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}} \qquad (4.1)$$

where $\underline{\mathbf{R}}\ (I \times J \times K)$ is the instrument response matrix, $\mathbf{x}_n$, $\mathbf{y}_n$, and $\mathbf{z}_n$ are the nth columns of the matrices $\mathbf{X}\ (I \times N)$, $\mathbf{Y}\ (J \times N)$, and $\mathbf{Z}\ (K \times N)$ containing the N pure component profiles in each dimension. $\underline{\mathbf{E}}\ (I \times J \times K)$ is the error matrix, e.g., noise. For GC x GC-TOFMS data the dimensions are the column 1 separations (**X**), the column 2 separations (**Y**), and the mass spectra (**Z**). The

trilinear model as applied to GC x GC-TOFMS data is illustrated graphically in Figure 4.1. Data with the trilinear structure, like GC x GC-TOFMS data, is advantageous because signals which are not completely resolved by the instrument can often be mathematically resolved if there is some selectivity in each of the three dimensions. This mathematical resolution, or deconvolution, does not require peak shape assumptions or completely selective mass channels.

4.2.2 Trilinear Decomposition (TLD) and Parallel Factor Analysis (PARAFAC)

TLD and PARAFAC are chemometric techniques that have been well documented in the literature [Booksh et al. 1994; Bro 1997; Faber et al. 2003; Li and Gemperline 1993; Sánchez and Kowalski 1990]. TLD is an eigenvalue-based solution to the trilinear PARAFAC model [Faber 2002; Sánchez and Kowalski 1990]. The alternating least squares (ALS) based solution to the trilinear PARAFAC model has acquired the name PARAFAC and is the most popular method. PARAFAC deconvolution uses a starter solution, in these experiments the TLD results, followed by ALS to find a solution to the model [Bro 1997; Faber et al. 2003]. Other possible starter solutions include random values, random orthogonalized values and singular values. TLD initialization was found to be the fastest approach and gave the best results for GC x GC-TOFMS data. TLD is advantageous because it does not require a starter solution and because it is computationally fast. But for the data in this chapter, TLD followed by PARAFAC deconvolution gives better results than TLD alone. Superior deconvolution with PARAFAC agrees with the findings of other authors [Faber et al. 2003]. The strength of PARAFAC deconvolution

is mainly attributed to non-negative and unimodal constraints that are incorporated in to the PARAFAC deconvolution algorithm.

4.3 Experimental

An Agilent 6890 gas chromatograph equipped with an Agilent 7683 auto-injector (Agilent Technologies, Palo Alto, CA, USA) was modified to a high-temperature valve-based GC x GC by mounting the wetted portions of the high-speed six-port micro diaphragm valve (VICI, Valco Instruments Co. Inc., Houston, TX, USA) inside the oven and the remaining portions outside the oven, freely exposed to room air [Sinha et al. 2003a]. The second column was then connected to a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI, USA) via the heated transfer line. Additional details about the instrument have been recently reported [Sinha et al. 2003b; Synovec et al. 2003a]. A GC x GC-TOFMS analysis was performed on a complex environmental sample (#1) to demonstrate the separation and performance characteristics of the new configuration. A complicated region of environmental sample #1 was then analyzed using TLD and PARAFAC to investigate the peak deconvolution capabilities of these techniques combined with the GC x GC-TOFMS data.

The first column (column 1) of the GC x GC-TOFMS for the complex environmental sample analysis was a 60-m 250-μm i.d. capillary column with a 0.5-μm 5% diphenyl/95% dimethyl polysiloxane film (DB-5; J&W Scientific/Agilent Technologies). The second column (column 2) was a 3-m 180-μm i.d capillary column with a 0.05-μm 90% biscyanopropyl/10% phenylcyanopropyl film (RTX-2330; Restek Corp., Bellefonte, PA, USA). Helium was used as the carrier gas. Column 1 was operated with a constant flow of 1.0 ml/min. Column 2 was operated with a constant pressure of 20

psi. The injector set point was 275 °C and 2.0 µl injections of environmental sample #1 were splitless for 0.5 minutes. The oven was held at 40 °C for 0.5 minutes, ramped to 80 °C at 20 °C/min, ramped to 210 °C at 5 °C/min, then ramped to 230 °C at 20 °C/min and held for 4 minutes. The valve was equipped with a 5-µl sample loop and actuated every 2.5 seconds with a 60-ms injection pulse width. A novel stand-alone pulse generator that was designed and built in-house was used to control the valve actuation. This new hardware allows the user to set the pulse width, the period of the actuations, and the total duration of the analysis. This new development replaces the LabVIEW 6i (National Instruments, Austin, TX, USA) program and counter/timer board [Sinha et al. 2003b; Synovec et al. 2003a]. The mass spectrometer had a transfer line temperature of 250 °C and an ion source temperature of 200 °C. The filament bias voltage was -70 V and the detector voltage was -2000 V. All other MS parameters were set from the results of an automatic optimization sequence controlled by the LECO software using perfluorotributylamine (PFTBA) as the standard. Data were collected from *m/z* 40 to 300 at a nominal rate of 5 kHz and averaged to 50 full spectra/second by the LECO software. Data were then exported as a comma separated value (.csv) file and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing. The algorithm for TLD was from the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA) and was chosen because of the advantageous ordering of the three dimensions of the matrix prior to analysis. The PARAFAC algorithm was from the N-way Toolbox 2.01 [Andersson and Bro 2000]. Chlorobenzene was used for a standard addition analysis such that the added amount was 4.3 µg/ml in environmental sample #1 (Alfa Products, Thiokol/Ventron Division, Danover, MA, USA).

4.4 Results and Discussion

A total ion current (TIC) chromatogram of complex environmental sample #1 is shown in Figure 4.2. Even though the overall separation achieved on environmental sample #1 was good, there were instances where components are not fully resolved with the two-dimensional chromatography. This is a common occurrence for highly complex samples. One such instance in environmental sample #1 is shown in Figure 4.3A. The sub-region defined by the inset box in Figure 4.3A is depicted in Figure 4.3B as a 3D image. In this region, there are two main components as well as some background and baseline signals. The feasibility of using TLD and PARAFAC with data from this instrument was investigated using these overlapped components as analytes and the region shown in Figure 4.3B as the known elution region of the analytes.

Figures 4.4A and B depict the column 1 and 2 peak profile results of TLD deconvolution, respectively. The TLD model was developed with four components (i.e., factors), but only three are shown for clarity, the fourth being mostly baseline offset which was not subtracted prior to deconvolution. Column 1 and 2 pure component profiles are reasonably good, although there is some deviation from conventional chromatographic peak shape in the column 2 profile of component labeled MPA. This is indicative that the TLD model is not able to completely deconvolute component MPA from the sloping edge of the adjacent background signal labeled "Bkgd" because of insufficient selectivity on column 1 (Figure 4.4B). The background signal is comprised of additional, unknown interferences. TLD also provides the deconvoluted mass spectra. Based on the deconvoluted mass spectral results, the sample sub-region was found to contain 1-methoxy-2-propyl acetate (MPA) and chlorobenzene (ClBz), as well as some unknown background

signals. There are multiple factors to evaluate the quality of a spectral match in the NIST MS search program. The match factor evaluates how closely the target spectrum and the database spectrum correlate. It is calculated based on the inner product of the two spectra, with lower *m/z* peaks having less weight than higher *m/z* peaks. Values are reported on a scale from 1-1000 where a perfect match is 1000. The reverse match factor ignores impurity peaks in the experimental spectrum, that is, peaks that are not present in the library spectrum. This is also reported on a scale of 1-1000. The probability of the unknown spectrum arising from the same compound that generated the library spectrum is listed on a scale from 1-100. The mass spectrum of component MPA was matched to a spectrum of 1-methoxy-2-propyl acetate in the NIST Library with a match factor of 904, a reverse match factor of 939, and a probability of 92.2 (Figure 4.5A and B). The mass spectrum of component ClBz was matched to a spectrum of chlorobenzene in the NIST Library with a match factor of 933, a reverse match factor of 937, and a probability of 98.4 (Figure 4.5C and D). The identity of component ClBz was confirmed with a standard addition of chlorobenzene into the original sample. The reverse match factor is higher than the match factor for component MPA (1-methoxy-2-propyl acetate) indicating that there are some *m/z* values in the deconvoluted spectrum that are not present in the library spectrum. This is most likely due to the component "Bkgd" contributing to the deconvoluted mass spectrum for MPA due to the low resolution on column 1. The spectral match is still quite high and the identification is accurate. The chromatographic profiles on column 2, however, are not ideal. Even though the trilinear chemometric methods do not rely upon peak shape information for the deconvolution, the analyst often desires to obtain deconvoluted peak

shapes of superior quality. Thus, reasonable chromatographic peak shapes should be strived for.

PARAFAC was employed on this same data in an effort to improve upon the peak deconvolution results obtained by TLD. The model was initiated with the TLD results shown in Figures 4.4 and 4.5. Non-negativity constraints were imposed on all dimensions. The column 1 and column 2 peak shapes are improved by PARAFAC over the TLD results, which should enable more accurate quantification (Figure 4.6A and B). Likewise, the deconvoluted spectra for analytes MPA and ClBz are shown in Figure 4.7A and B. The mass spectrum of component MPA was matched to a spectrum of 1-methoxy-2-propyl acetate (Figure 4.5B) in the NIST Library with a match factor of 879, a reverse match factor of 905, and a probability of 88.7. The match factor for MPA with the PARAFAC is slightly less than that of TLD, but the difference is not statistically significant. By comparing both deconvoluted spectra of MPA (Figures 4.5A and 4.7A) with the library spectrum of 1-methoxy-2-propyl acetate (Figure 4.5B), it can be seen that the ratios of the higher masses of the TLD result correspond better with the library spectrum than those for the PARAFAC result. This could be the cause of the slight decrease in match quality since the higher *m/z* ratios have more weight in calculating the match factor. The mass spectrum of component ClBz was matched to a spectrum of chlorobenzene in the NIST Library (Figure 4.5D) with a match factor of 944, a reverse match factor of 948, and a probability of 98.5, indicating a very good identity match.

The above example indicates the GC x GC-TOFMS data is compatible with chemometric calibration techniques like PARAFAC and TLD that call for trilinear data structure. Even though some of the ions appear selective for each analyte, e.g., *m/z* 112 for chlorobenzene and *m/z* 43 for 1-methoxy-2-

propyl acetate, the TLD and PARAFAC algorithms do not require selective ions. TLD and PARAFAC are able to deconvolute pure analyte profiles from complex mixtures as long as there is at least partial selectivity in each of the three dimensions. In general, very complex samples may contain overlapped constituents that do not have a selective ion, e.g., isomers. A study of complex samples with isomers is presented in Chapter 5 to investigate this issue more thoroughly. Furthermore, unlike deconvolution methods developed for GC x GC with flame ionization detection, TOFMS detection provides additional structure to the data, which allows for deconvolution using trilinear-based methods without the comparison of data sets. This substantially relaxes retention time reproducibility requirements and makes for a more reliable deconvolution. The selectivity of TOFMS substantially reduces the probability of unresolved signals as well.

From this initial study, a four-step procedure has been developed to deconvolute, identify and quantify analytes of interest that are not fully resolved nor have a fully selective mass channel. First, two data files are collected: A "sample" and a "sample + standard addition," in which quantitative amounts of all the analytes of interest are spiked into the standard addition. TLD followed by PARAFAC is performed on the region around each analyte of interest, which will give individual peak profiles and mass spectra for both data sets. The analytes in the sample data set are identified by comparing their deconvoluted mass spectra to those of the standard addition sample or to a MS library. Comparison to the standard addition sample can improve the quality of matches because the reference spectrum (i.e., standard addition sample) is obtained on the same instrument as the sample spectrum as opposed to the NIST library spectra that are obtained on a number of different instruments resulting in different

fragmentation ion ratios. Quantification is then achieved by reconstructing the signal of the analytes of interest in both data sets then applying signal integration and the usual mathematical techniques for quantification via standard addition. The fact that the deconvolution is performed separately on both the sample and the standard addition loosens the retention time alignment requirements, thus simplifying and improving the quantification process significantly. This analysis procedure was followed for the identification and quantification of chlorobenzene in environmental sample #1. A chlorobenzene standard was spiked into the sample as a standard addition at the level of 4.3 μg/ml. Deconvolution was performed on a region of the standard addition data set containing the spiked standard. Reconstruction of the chlorobenzene peak and summing all of the mass channels to generate a TIC chromatogram was achieved for both the "sample" and "sample + standard addition sample." The volumes for the sample and standard addition peaks calculated from the respective TIC chromatograms indicated that the original concentration of chlorobenzene in environmental sample #1 was 1.4 μg/ml. Thus quantification via standard addition without retention time alignment was successful using TLD-initiated PARAFAC. The precision and accuracy for TLD and PARAFAC are consistent with second order methods such as GRAM and indicate TLD-initiated PARAFAC appears to perform better at lower chromatographic resolution than does GRAM. These attributes of TLD-initiated PARAFAC will be studied in more detail in future work.

4.5 Conclusions

It was demonstrated that the GC x GC-TOFMS trilinear data structure is compatible with chemometric calibration techniques such as TLD and

PARAFAC. Successful deconvolution using PARAFAC initiated by TLD, followed by analyte identification was achieved on overlapped constituents in a complex environmental sample, requiring only a single data set for qualitative analysis. Analytes of interest in a mixture can be identified and quantified using a standard addition method combined with TLD-initiated PARAFAC, which would eliminate the need for peak shape predictions and retention time alignment between the sample and the standard prior to deconvolution. TLD-initiated PARAFAC also eliminates the need for fully selective mass channel ions for deconvolution; however, some selectivity is required in each dimension for the algorithm to be successful. Future work will be aimed to probe the limits of algorithm applicability with regard to quantitation, precision, and accuracy of the results. In addition, future work to automate this process for regions of interest will be addressed.

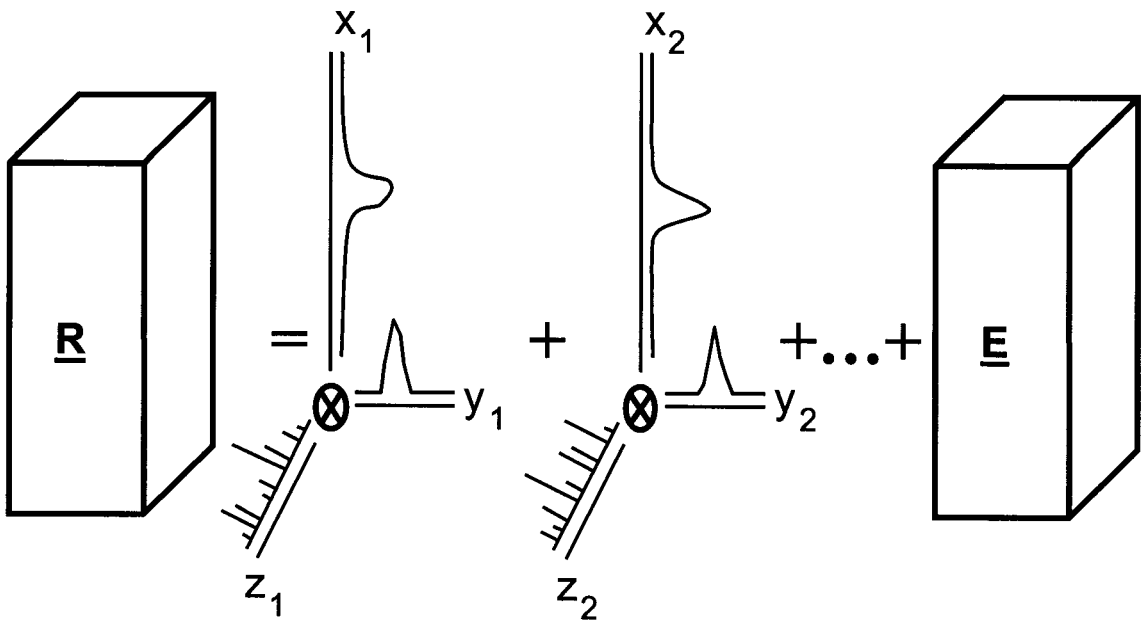


Figure 4.1. Illustration of the trilinear data structure of GC x GC-TOFMS data. For the instrument response $\underline{\mathbf{R}}$ there are distinctive profiles in both chromatographic dimensions ($\mathbf{x}_n$ and $\mathbf{y}_n$) and a unique spectrum ($\mathbf{z}_n$) for each constituent in a data matrix. The error (e.g., noise) is denoted $\underline{\mathbf{E}}$. This is mathematically described as $\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}}$.

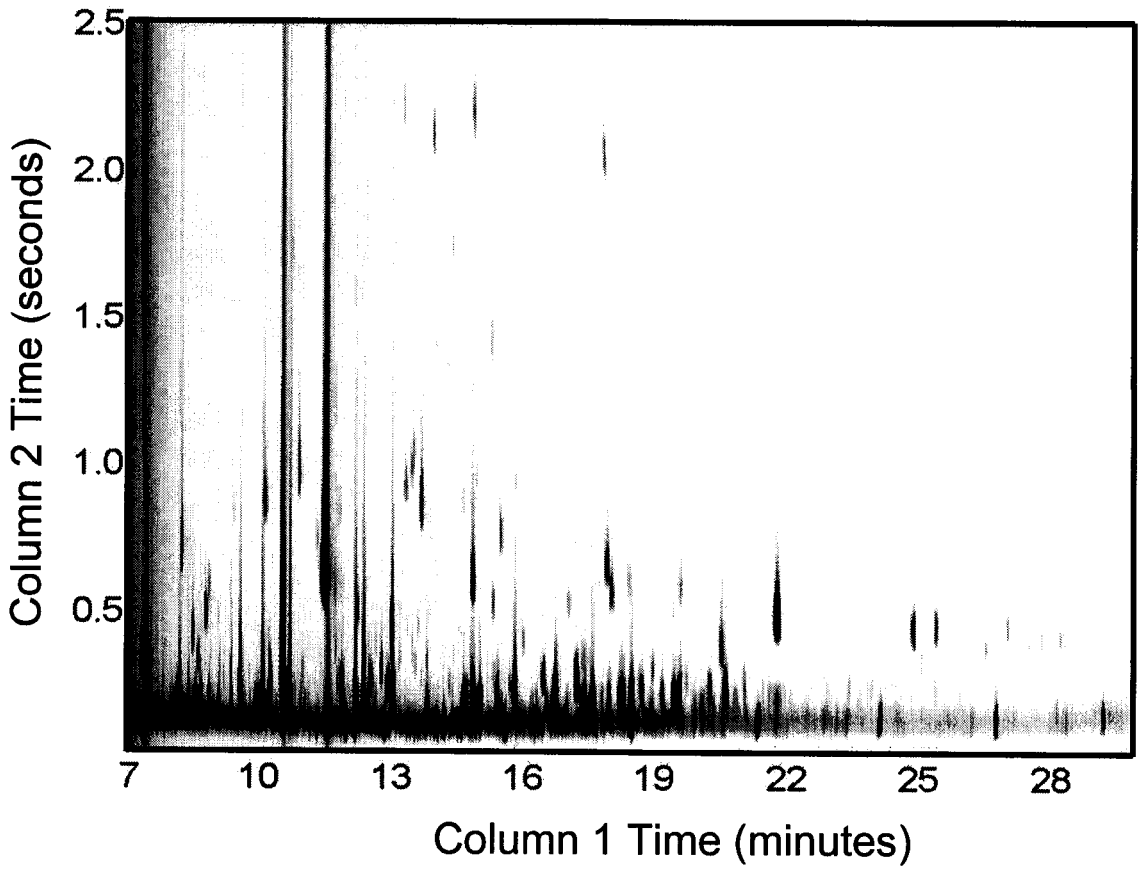


Figure 4.2. TIC chromatogram of complex environmental sample #1.

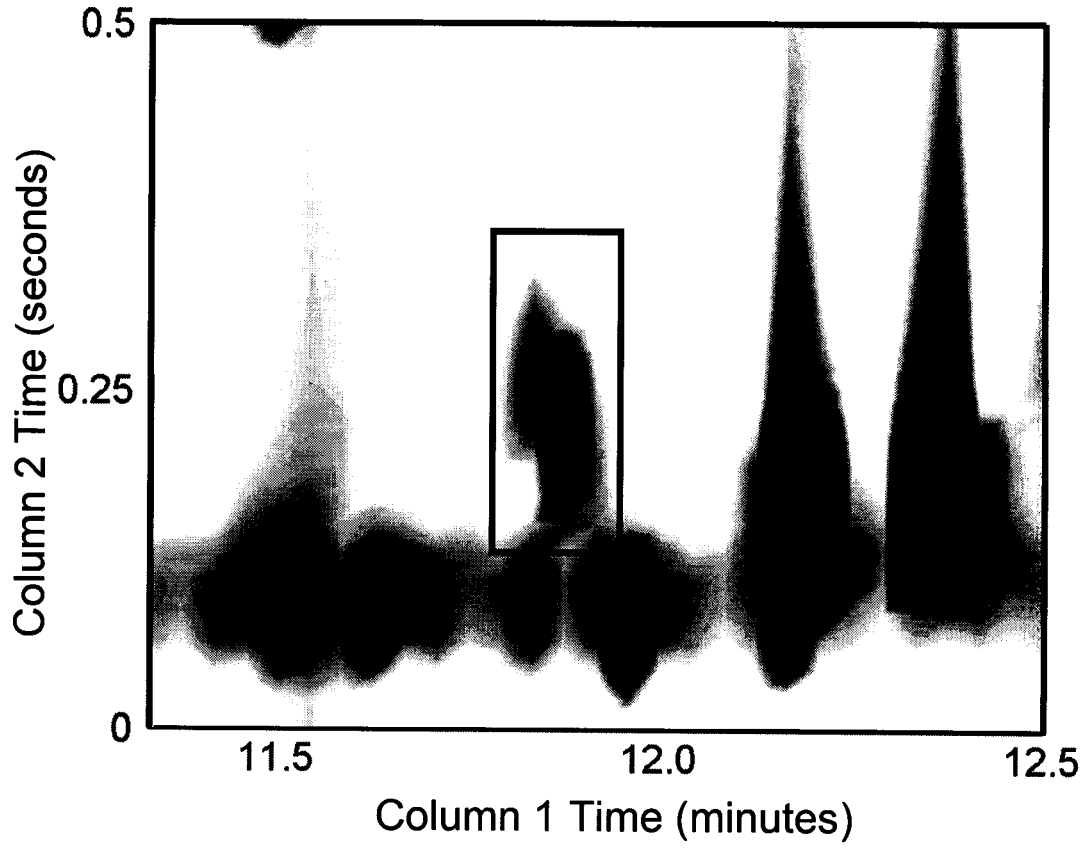


Figure 4.3A. Region of the TIC of environmental sample #1. This region contains overlapped peaks as indicated by overlaid box.

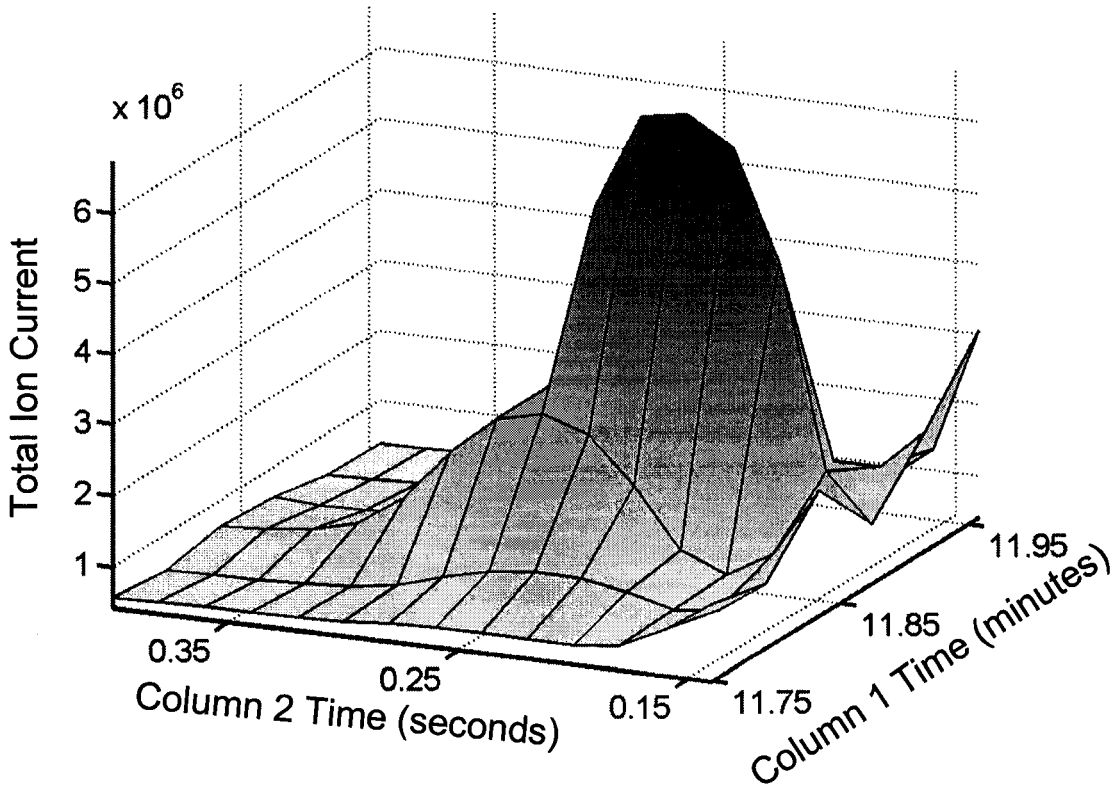


Figure 4.3B. 3D image of the sub-region in the box outlined in Figure 4.3A.

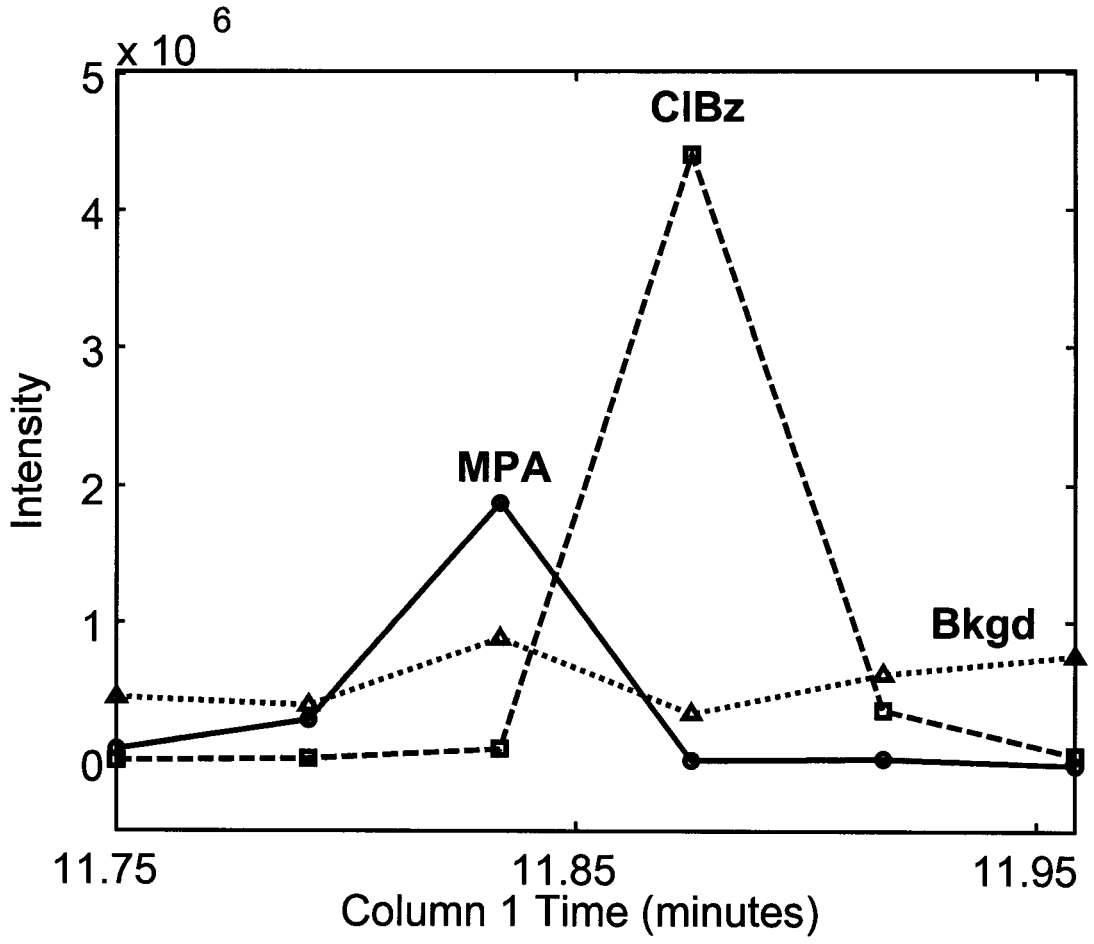


Figure 4.4A. TLD deconvoluted column 1 pure component profiles. Results are for the deconvolution of two chemical components in environmental sample #1. Three of the four components are shown: **MPA,C lBz** and background interferents (**Bkgd**).

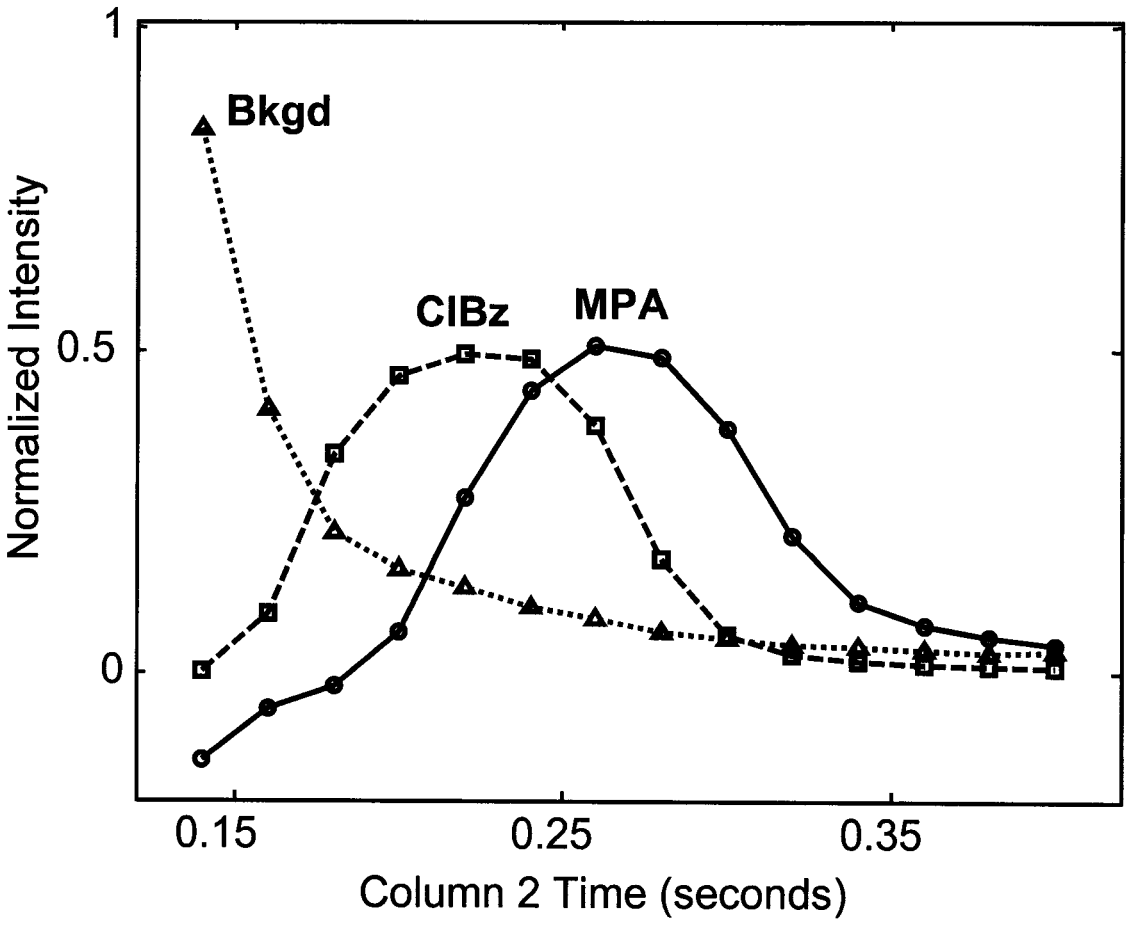


Figure 4.4B. TLD deconvoluted column 2 pure component profiles. **MPA** and **ClBz** are identified by the deconvoluted mass spectra in Figure 4.5. The fourth component modeled was "baseline", omitted for clarity.

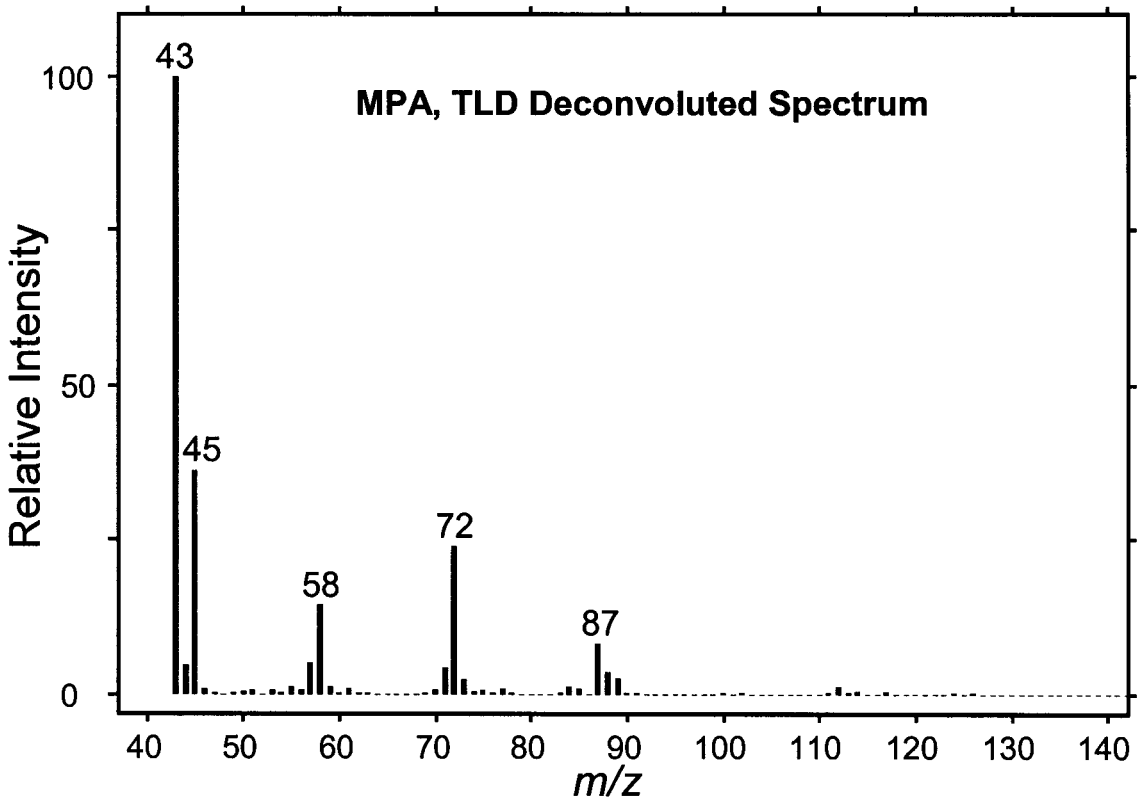


Figure 4.5A. TLD deconvoluted mass spectrum for analyte **MPA**. Results are from the deconvolution of two components in environmental sample #1 whose chromatographic profiles appear in Figures 4.4A and 4.4B. **MPA** is 1-methoxy-2-propyl acetate.

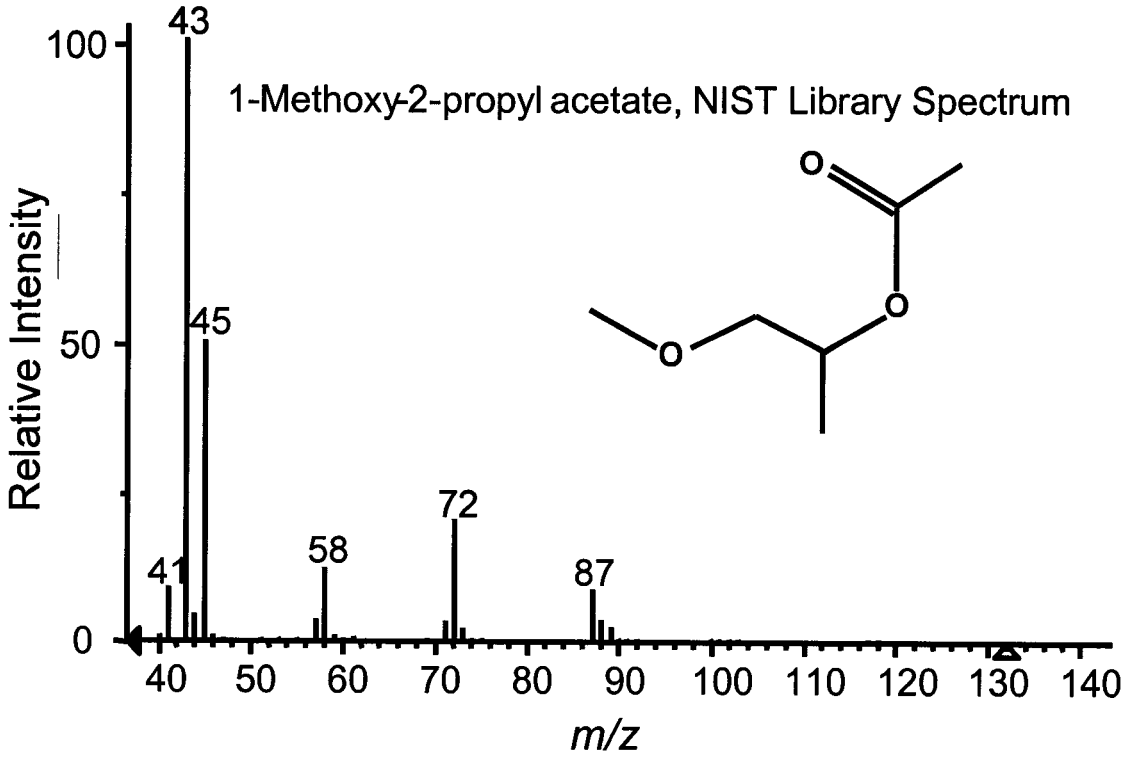


Figure 4.5B. NIST Library mass spectrum of **MPA**.

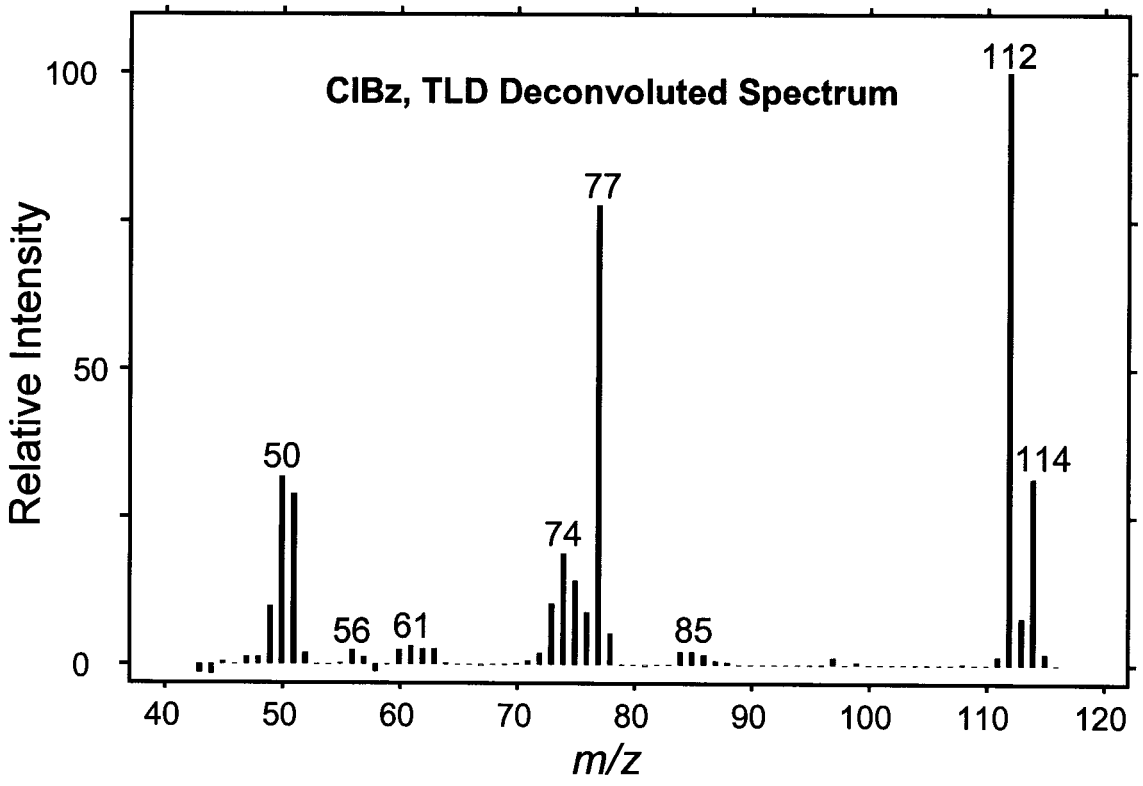


Figure 4.5C. TLD deconvoluted mass spectrum for analyte **ClBz**. Results are from the deconvolution of two components in environmental sample #1 whose chromatographic profiles appear in Figures 4.4A and 4.4B. **ClBz** is chlorobenzene.

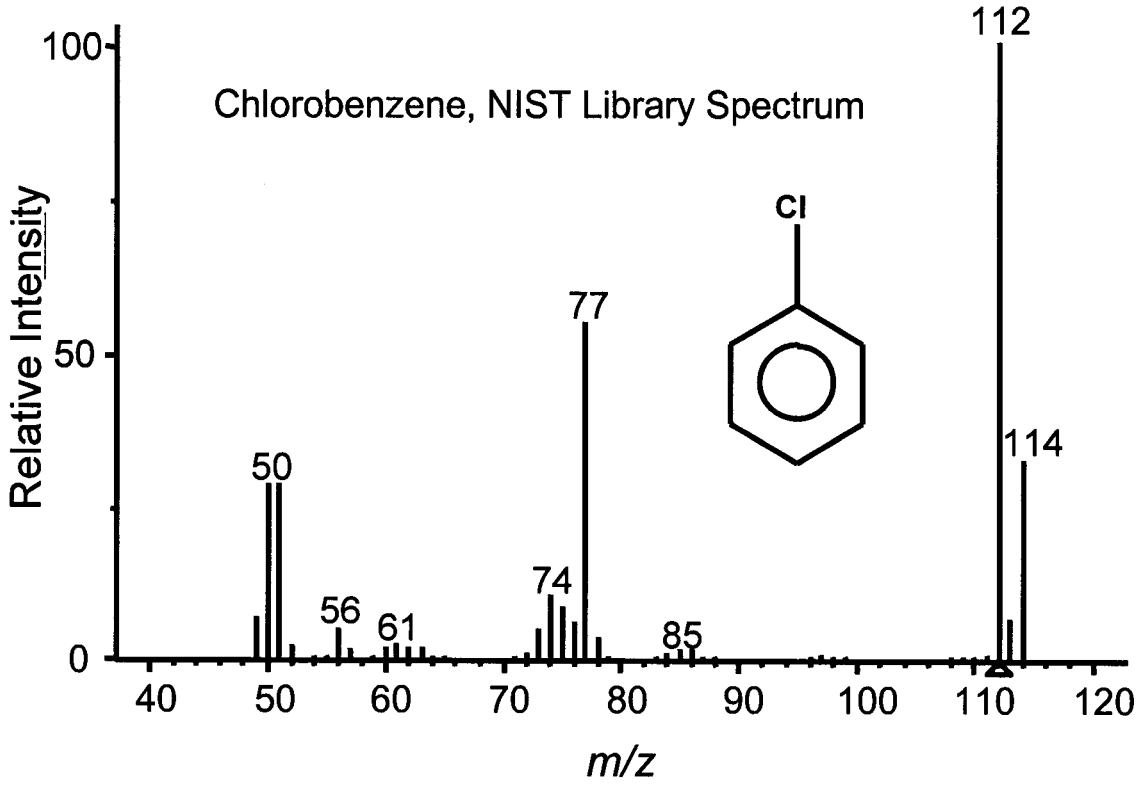


Figure 4.5D. NIST Library Mass spectrum of analyte **ClBz**. Further, confirmation of the identity of analyte **ClBz** as chlorobenzene was achieved through standard addition.

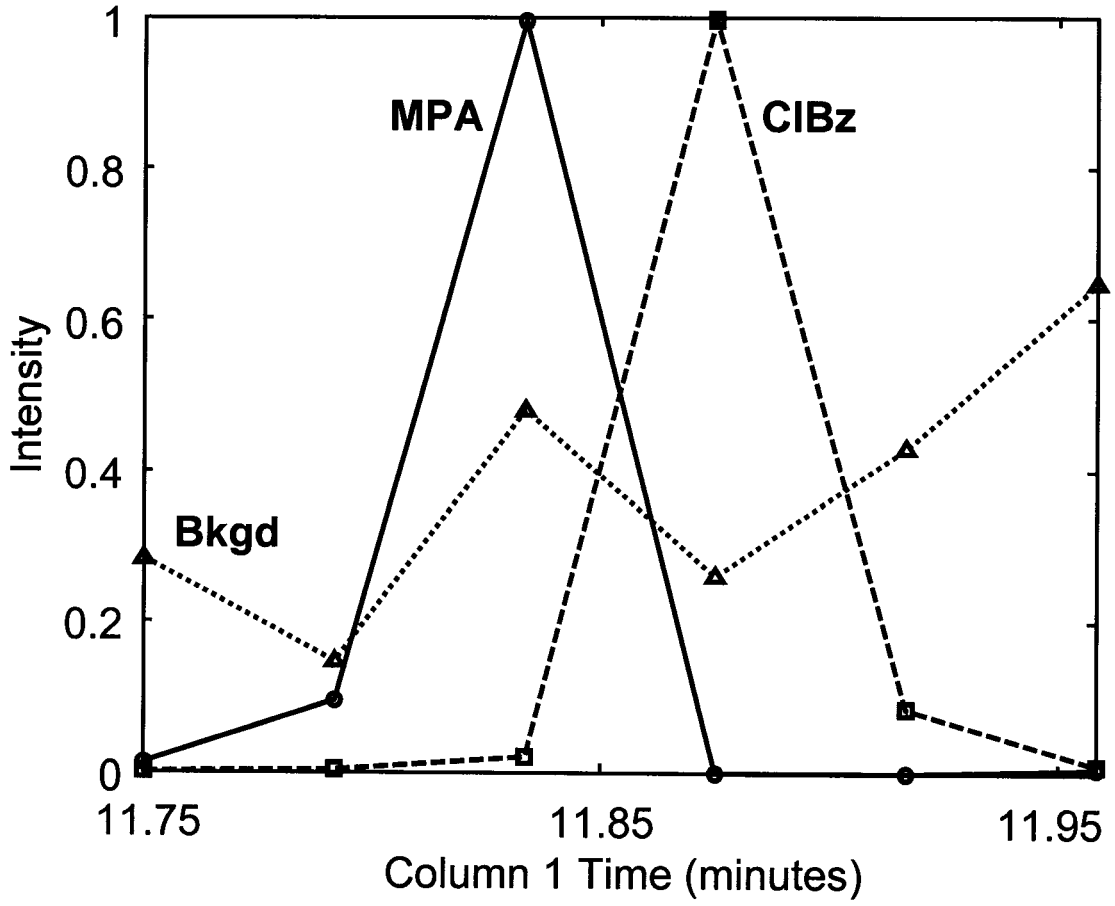


Figure 4.6A. PARAFAC deconvoluted column 1 pure component profiles. Results are from the deconvolution for the overlapped region in environmental sample #1 using TLD results for initiation and using non-negativity constraints in all dimensions. As in Figure 4.4A, three of the four components are shown for clarity.

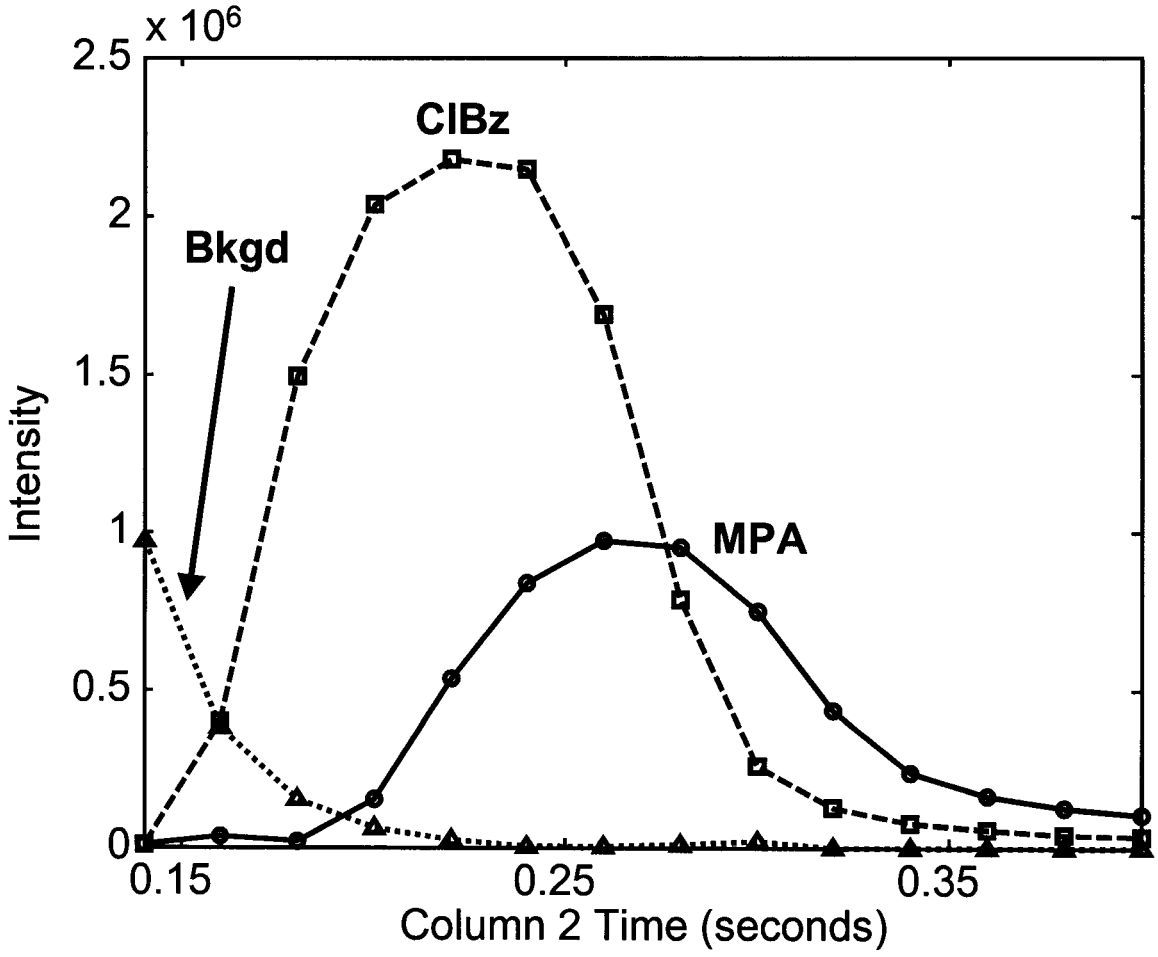


Figure 4.6B. PARAFAC deconvoluted column 2 pure component profiles. Results are from the deconvolution for the overlapped region in environmental sample #1 using TLD results for initiation and using non-negativity constraints in all dimensions. As in Figure 4.4B, three of the four components are shown for clarity.

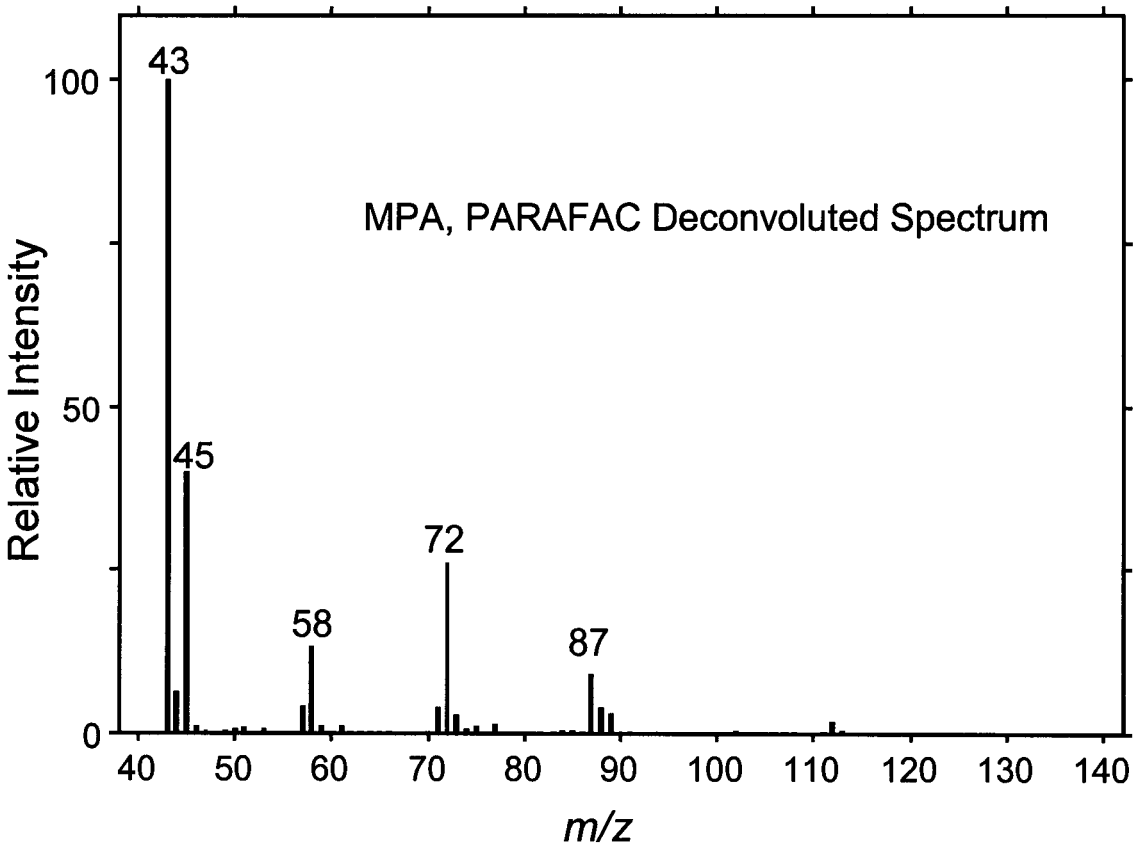


Figure 4.7A. PARAFAC deconvoluted mass spectrum for analyte **MPA**. Results are for overlapped region in environmental sample #1 using TLD results for initiation and using non-negativity constraints in all dimensions. MPA is 1-methoxy-2-propyl acetate. NIST Library MS search resulted in the same chemical identity as obtained for TLD results (Figure 4.5B).

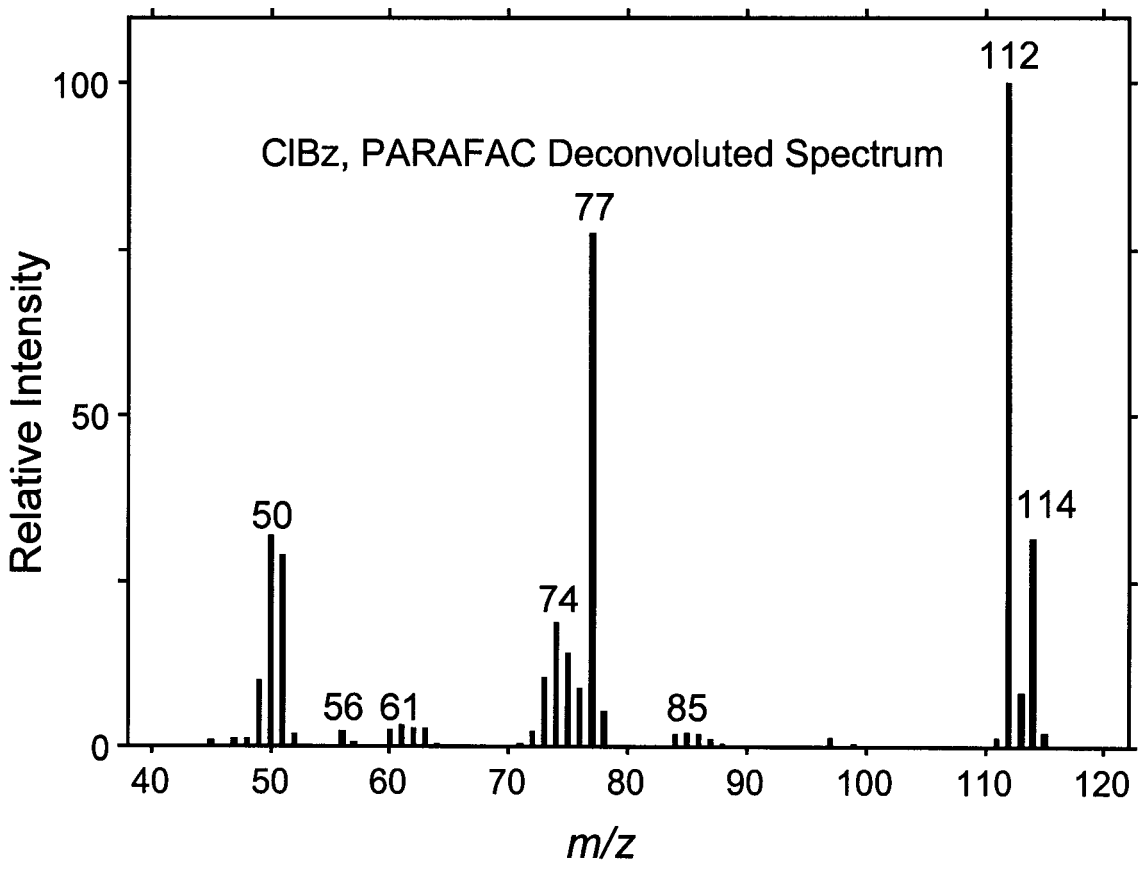


Figure 4.7B. PARAFAC deconvoluted mass spectrum for analyte **ClBz**. Results are for overlapped region in environmental sample #1 using TLD results for initiation and using non-negativity constraints in all dimensions. ClBz is chlorobenzene. NIST Library MS search resulted in the same chemical identity as obtained for TLD results (Figure 4.5D).

CHAPTER 5. Multivariate Selectivity as a Metric for Evaluating GC x GC - TOFMS Data Subjected to Chemometric Peak Deconvolution*

5.1 Introduction

GC x GC-TOFMS is quickly becoming a popular area of research [Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; Dimandja 2003; Korytar et al. 2003; Lu et al. 2003; Shellie et al. 2001b; Sinha et al. 2003b; van Deursen et al. 2000], fueled by tremendous selectivity and sensitivity for the identification and analysis of components in complex mixtures. Along with providing selectivity and sensitivity, GC x GC-TOFMS is an instrument capable of generating trilinear data, thus broadening the opportunity to utilize state-of-the-art chemometric techniques for signal deconvolution (i.e., mathematical resolution). The trilinear data structure of GC x GC-TOFMS can be an advantageous attribute for mathematically resolving overlapped signals. Certain conditions need to be met in order for data to be trilinear: "the response in [all] domains of the instrument arising from a species should be unique, consistent, and independent of the presence of other species" [Lin et al. 1994]. Trilinear data structure is the key to chemometric deconvolution techniques like the generalized rank annihilation method (GRAM), trilinear decomposition (TLD) and parallel factor analysis (PARAFAC) [Booksh et al. 1994].

Currently available peak deconvolution methods used for GC x GC-TOFMS data analysis are typically extensions of those that are already used

---

* Large portions of this Chapter have been previously published [Sinha et al. 2004b].

for GC-MS and do not utilize the trilinear data structure. These deconvolution methods reduce GC x GC-TOFMS data into a series of column 2 separation dimension GC-MS data. Each column 2 separation is analyzed independently, then recombined along the column 1 separation dimension for the final assessment. These methods essentially do not utilize the column 1 separation dimension during deconvolution of overlapped peaks.

With GC x GC combined with non-spectrometric detection like flame ionization detection, calibration methods such as GRAM are able to deconvolute chromatographic signals using two data sets (standard and sample) where the analytes of interest vary in concentration between the two data sets [Bruckner et al. 1998; Fraga et al. 2000a; Fraga et al. 2000b; Fraga et al. 2001a; Prazen et al. 1999a; Synovec et al. 2003b]. These techniques stack multiple bilinear data sets to create a trilinear data structure, which can then be deconvoluted. Using trilinear data, such as GC x GC-TOFMS data, it is possible to deconvolute individual components from a group of partially overlapped components using a data set from only one sample. Thus, analytes of interest can be identified in a complex mixture when there is only partial selectivity in both chromatographic dimensions and the mass spectral dimension, unlike currently available methods that rely upon selective ion deconvolution approaches. The ability to deconvolute a single data set from partially resolved signals into the fully resolved signals is known as the third-order advantage [Booksh and Kowalski 1994]. In the case of chromatographic data, third-order data is also advantageous because it relaxes the requirements for sample-to-sample retention time precision, which is a critical issue for other methods based on the trilinear data model such as GRAM [Fraga et al. 2001b], thus essentially eliminating the need for retention time alignment prior to analyte deconvolution.

Previously, an initial report demonstrated the use of TLD and PARAFAC for GC x GC-TOFMS data on an environmental sample [Sinha et al. 2004a]. Here the previous work is built upon by studying the effects of multivariate selectivity (chromatographic resolution and mass spectral similarity) on GC x GC-TOFMS peak deconvolution methods based on the trilinear data structure. The chemometric methodology will employ TLD to initiate PARAFAC. An investigation is presented with four butyl benzene isomers (sec-butyl, iso-butyl, tert-butyl and n-butyl benzenes) that have similar spectra with no selective major ions. In order to investigate TLD-initiated PARAFAC deconvolution of peaks with similar spectra at varying chromatographic resolution, three replicate data sets were collected by GC x GC-TOFMS for each of four isomers (sec-butyl, iso-butyl, tert-butyl and n-butyl benzenes) at a concentration of 3% volume/volume in hexane. These individual data sets of the isomers were then added together to create "constructed" samples. These constructed samples have the appearance of a separation of a mixture of the four isomers. Although these isomers can be separated by GC x GC, they are used here as an investigative tool to probe the practical limits of third-order chemometric techniques for peak deconvolution. The major benefits of this approach are that the resolution (i.e., effective chromatographic selectivity) of the isomers in each chromatographic dimension can be readily altered, the true expected deconvolution results are known, and noise in the data is real. All analyses were performed in triplicate and after baseline correction. The chemometric models were calculated with four components specified. The PARAFAC analysis also incorporated non-negative constraints in all dimensions and unimodal constraints for both chromatographic dimensions. In addition, a derivatized metabolic plant extract from Huilmo (*Sisyrinchium striatum*) containing a selected region of

overlapped peaks is also analyzed to demonstrate the technique on a novel, complex natural sample. Future studies involving metabolites in complex samples will explore the application of pattern recognition methods that take advantage of the trilinear structure to GC x GC-TOFMS data [Johnson and Synovec 2002; Prazen et al. 2001].

5.2 Theory

5.2.1 Trilinear Data

Mathematically, the PARAFAC model is described as

$$\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}} \qquad (5.1)$$

where $\underline{\mathbf{R}}$ $(I \times J \times K)$ is the instrument response matrix, $\mathbf{x}_n$, $\mathbf{y}_n$, and $\mathbf{z}_n$ are the nth columns of the matrices $\mathbf{X}$ $(I \times N)$, $\mathbf{Y}$ $(J \times N)$, and $\mathbf{Z}$ $(K \times N)$ containing the N pure component profiles in each dimension. $\underline{\mathbf{E}}$ $(I \times J \times K)$ is the error matrix, e.g., noise, and $\otimes$ denotes the mathematical function for the outer (or cross) product. For GC x GC-TOFMS data the dimensions are the column 1 separation space (**X**), the column 2 separation space (**Y**), and the mass spectral dimension (**Z**). Figure 5.1 contains a graphical representation of the trilinear model as applied to GC x GC-TOFMS data. Data with the trilinear structure, like GC x GC-TOFMS data, is gainful because signals which are not fully resolved by the instrument can normally be mathematically resolved if there is at least some selectivity in each of the three dimensions. This mathematical resolution, or deconvolution, does not entail peak shape assumptions or fully selective mass channels.

5.2.2 Multivariate Selectivity and Net Analyte Signal (nas)

When quantifying the selectivity of GC x GC-TOFMS data in real examples, chromatographic resolution alone often does not fully describe the selectivity of GC x GC-TOFMS data. Chromatographic resolution describes the selectivity of a peak when it is overlapped with a single interfering peak; however, when there are multiple interfering peaks, resolution cannot fully describe the information. Also, the selectivity contained in the spectra from hyphenated techniques, like GC x GC-TOFMS, can play a critical role in determining the information content of overlapped peaks, but this is not accounted for in the resolution equation. For example, two peaks that have low-resolution on both chromatographic dimensions could have nearly selective mass channels and thus contain a great deal of analytical information. To deal with this issue, chemometricians have introduced a metric known as multivariate selectivity [Booksh and Kowalski 1994; Faber et al. 1997; Lorber et al. 1997]. Multivariate selectivity is the degree of overlap of an analyte's signal with signals from different sources. Multivariate selectivity ranges between zero (complete overlap) and unity (no overlap). A plot comparing multivariate selectivity to two-dimensional chromatographic resolution is shown in Figure 5.2. For low-resolution cases the small changes in resolution change the selectivity drastically. Once two peaks are resolved by about 0.75, further resolution does not substantially increase the multivariate selectivity. These trends are analogous to the results from chromatographic deconvolution. For chromatographic analysis, multivariate selectivity can describe the information content of peaks overlapped with multiple interference peaks and can be extended to describe data from

hyphenated chromatographic separation and spectrometric data. We have found that multivariate selectivity is strongly correlated with the quality of results when techniques that are based on the trilinear data model are used for deconvolution and quantification. Like chromatographic resolution, multivariate selectivity is independent of both concentration and detector sensitivity, and thus can be combined with signal-to-noise or sensitivity for a complete description of data. Multivariate selectivity (SEL) is defined as the ratio of the net analyte signal (nas) to the total analyte signal, as shown in Eq. (5.2):

$$\mathrm{SEL} = \frac{\|\mathrm{nas}(\mathbf{s})\|}{\|\mathbf{s}\|} \qquad (5.2)$$

where $\|\mathbf{s}\|$ denotes the norm of the signal ($\mathbf{s}$). In geometrical terms, net analyte signal, nas($\mathbf{s}$), is defined as the portion of an analyte's signal that is orthogonal to the signals of the other components in the sample matrix. The net analyte signal of a vector of data, such as a chromatogram or spectrum, is described mathematically as

$$\mathrm{nas}(\mathbf{s}) = (\mathbf{I} - \mathbf{X} * \mathbf{X}^{+})\mathbf{s} \qquad (5.3)$$

where $\mathbf{X}$ is a matrix of chromatograms or spectra of all components in the sample except that of the analyte. $\mathbf{X}^{+}$ denotes the pseudoinverse of $\mathbf{X}$, and $\mathbf{I}$ is the identity matrix with the same dimensions as $\mathbf{X}$. For the description of three-dimensional signals like GC x GC-TOFMS the selectivity on each dimension is multiplied to obtain the three-dimensional selectivity. This is the reason that some selectivity is required on each of the chromatographic and spectral dimensions to do mathematical resolution using trilinear techniques. It is important to remember that multivariate selectivity can be misleading in multidimensional data when there is nearly complete selectivity in some of the dimensions but not all of the dimensions. In this case, the acute analyst

would only need to use the selective dimensions for quantitation and identification, thus arriving at a multivariate selectivity of nearly one, by discarding the less selective dimensions. In this report, multivariate selectivity is used as a metric to describe the analytical information content of peaks overlapped on both of the chromatographic dimensions and the mass spectral dimension.

5.2.3 TLD and PARAFAC Deconvolution

TLD and PARAFAC are multivariate techniques for peak deconvolution and calibration that have been well documented in the literature [Booksh et al. 1994; Bro 1997; Faber et al. 2003; Li and Gemperline 1993; Sánchez and Kowalski 1990]. TLD is an eigenvalue-based solution to the trilinear PARAFAC model [Faber 2002; Sánchez and Kowalski 1990]. PARAFAC is the alternating least squares (ALS) based solution to the trilinear PARAFAC model [Bro 1997; Faber et al. 2003]. PARAFAC requires a starter solution, in these experiments the TLD results, to model the data. Other feasible starter solutions include random values, random orthogonalized values and singular values. TLD initialization for PARAFAC was found to be the fastest and gave the best results for GC x GC-TOFMS data. TLD is advantageous in that it does not require a starter solution and because it is computationally fast, but for the results presented in this report, TLD-initiated PARAFAC provided better results than TLD alone. The advantage of PARAFAC deconvolution can be attributed predominantly to non-negative and unimodal constraints that are incorporated in to the PARAFAC deconvolution algorithm. Constraints cannot be incorporated into the TLD algorithm. Superior deconvolution with PARAFAC agrees with the results of other authors [Faber et al. 2003]. For brevity, only TLD-initiated PARAFAC

results are presented here, and TLD results are omitted. PARAFAC requires some selectivity on each of the chromatographic and spectral dimensions, but avoids the requirement of selective ions.

5.3 Experimental

5.3.1 Butyl Benzene Isomers

An Agilent 6890 Gas Chromatograph (Agilent Technologies, Palo Alto, CA, USA) was modified to a valve-based GC x GC by mounting the portions of the valve (VICI, Valco Instruments Co. Inc., Houston, TX, USA) which are exposed to sample inside the oven and the remaining portions outside the oven uncovered and exposed to room air [Sinha et al. 2003a]. The instrument was equipped with an Agilent 7683 auto-injector (Agilent Technologies, Palo Alto, CA, USA). Column 2 of the GC x GC was connected to a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI, USA) via the heated transfer line. A diagram of the instrument is contained in a recent publication [Sinha et al. 2003b]. A set of four butyl benzene isomers with similar spectra was used to investigate the effect of changing selectivity on the TLD and PARAFAC results by changing the chromatographic resolution in the GC x GC separation space.

Four butyl benzene isomers (sec-butyl, iso-butyl, tert-butyl and n-butyl benzenes) (99% purity; Sigma-Aldrich Corp., St. Louis, MO, USA) were used to create 3% vol/vol solutions of each individual isomer in hexane (96% n-hexane; J.T. Baker, Phillipsburg, NJ, USA). Three replicate data sets were obtained for each solution. Column 1 of the GC x GC-TOFMS for the butyl benzene isomers was a 20-m 180-μm i.d. capillary column with a 0.18-μm 5% diphenyl/95% dimethyl polysiloxane film (RTX-5; Restek Corp.). Column 2 was a 3-m 180-μm i.d capillary column with a 0.05-μm 90%

biscyanopropyl/10% phenylcyanopropyl film (RTX-2330; Restek Corp.). Ultra high purity helium was used as the carrier gas. Column 1 was operated with a constant pressure of 20 psi (138 kPa). Column 2 was operated with a constant pressure of 2 psi (13.8 kPa). The injector set point was 275 °C and 0.2-µl injections of 3% vol/vol solutions of each isomer in hexane were split 50:1. The oven was operated isothermally at 100 °C for 4 minutes. The valve was equipped with a 5-µl sample loop and actuated at a rate of 1 Hz with a 20 ms injection pulse width. A stand-alone pulse generator was used to control the valve actuation [Sinha et al. 2003b]. The mass spectrometer had a transfer line temperature of 250 °C and an ion source temperature of 200 °C. The filament bias voltage was -70 V and the detector voltage was -1500 V. All other TOFMS parameters were set from the results of an automated optimization sequence controlled by the LECO software using perfluorotributylamine (PFTBA) as the standard. Data were collected from *m/z* 30 to 150 at a nominal rate of 5 kHz and averaged to 100 full spectra/second by the LECO software. Data were then exported as a comma separated value (.csv) file and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing. The algorithm for TLD was from the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA) and was selected for the advantageous sequencing of the three dimensions of the matrix prior to analysis. The PARAFAC algorithm was from the N-way Toolbox 2.01 [Andersson and Bro 2000]. Mass spectral similarity searches were performed with NIST MS Search 2.0 (NIST/EPA/NIH Mass Spectral Library; NIST 98). Baseline correction was done at each mass with a linear correction along the second column time dimension over the region subjected to deconvolution.

5.3.2 Huilmo (*Sisyrinchium striatum*) Metabolite Extracts

A natural plant metabolite sample of Huilmo (*Sisyrinchium striatum*) was also analyzed to demonstrate TLD-initiated PARAFAC on a natural sample. Prior to analysis, metabolites were extracted and derivatized via trimethylsilylation [Adams et al. 1999; Fiehn et al. 2000]. The analysis of the extracted sample was performed using a thermally modulated LECO Pegasus 4D GC x GC-TOFMS instrument (LECO Corporation). Column 1 was a 10-m 180-μm i.d. capillary column with a 0.18-μm 5% diphenyl/95% dimethyl polysiloxane (DB-5; J & W Scientific, Alltech, Deerfield, Il). Column 2 was a 2-m 100-μm i.d. capillary column with a 0.1-μm film of (50%-Phenyl)-methylpolysiloxane (DB-17; Alltech). These columns were joined using a mini union (Scientific Glass Engineering SGE, Austin, TX). Modulation, or delivery of column 1 effluent onto column 2 was performed using cryogenic modulation. Effluent from column 1 was concentrated at the head of column 2 during each column 2 separation. A "hot pulse," occurring when the cryogenic gas was switched off and the heated air jets (40 °C above the oven temperature) were turned on, was used to begin each new column 2 run. The hot pulse was 0.40 seconds in duration, and the total column 2 run time was 2 seconds. Ultra high purity helium (0.8 ml/min) was used as the carrier gas. 1 μl of derivatized sample was injected using a 25:1 split and a column 1 oven ramp beginning at 70 °C with a hold time of 5 minutes then increasing at 5 °C/min to 250 °C with a hold time at 250 °C of 5 minutes. Column 2 was held in a separate oven, which was held at a constant 40 °C higher than the column 1 temperature throughout the column 1 oven ramp. The first 5 minutes of each run was considered a solvent delay and no mass spectra were collected during that time. The transfer line was held at 260 °C and the ion source was

held at 200 °C. The detector voltage was -1600 V and the filament bias was -70 V. Mass spectra were collected from *m/z* 70 - 625 at a nominal rate of 5 kHz and averaged to 100 full spectra/second. Data were processed similarly to the butyl benzene isomers.

5.4 Results and Discussion

5.4.1 High-Resolution Butyl Benzenes Deconvolution

To verify that there was sufficient selectivity among the mass spectra of the four isomers, a benchmark "high-resolution" case was constructed such that the four isomers had at least unit resolution in the column 1 dimension and a resolution of 0.6 in the column 2 dimension (Figure 5.3A). It should be noted that since individual data sets were added together, the column 1 and column 2 times in the figure are for gauging relative retention times only and do not indicate real retention times. The column 1 and column 2 chromatographic peak profiles resulting from PARAFAC deconvolution of the high-resolution case yielded excellent quantitative results (Table 5.1) and excellent peak shapes (not shown for brevity). PARAFAC deconvolution of each dataset took about five minutes on a 1.6 GHz PC. The deconvoluted mass spectra (also not shown for brevity) were all matched to the appropriate library mass spectra with similarity match factors greater than 900 (where 999 is an exact match) and were generally within one standard deviation of the match factors obtained for the signals of the pure compounds prior to constructing the data sets. Standard deviations were determined using the three replicate data sets collected for each analyte. All matches listed in this report were the highest-ranking matches. The reproducibility and bias of the method for this high level of chromatographic resolution was studied by comparing the true (prior to addition) fraction of total peak volume for each

peak to the PARAFAC analyzed peak volume fraction for the deconvoluted peak. Normalization of the three deconvoluted replicate peak volume fractions to the mean of the true volume fractions was necessary to eliminate apparent biases due to run-to-run injection variability. The bias is defined as the difference between a normalized reconstructed peak volume fraction and the mean true peak volume fraction divided by the mean true peak volume fraction. The absolute values of the biases are all less than 3% for the benchmark "high-resolution" case, which suggests an acceptable level of accuracy (Table 5.1). The %RSD calculated and tabulated here is based upon the standard deviation of the normalized peak volume fractions divided by the mean of those fractions. The %RSDs for the benchmark case are all less than 1% indicating a high level of precision (Table 5.1).

5.4.2 Intermediate-Resolution Butyl Benzenes Deconvolution

In this case, the resolution between adjacent isomers on column 1 was nominally 0.25 and 0.2 on column 2 (Figure 5.3B). After baseline correction, the three replicate intermediate-resolution constructed data sets were analyzed. Similar to the high-resolution data sets, TLD-initiated PARAFAC of the intermediate-resolution data sets produced good deconvoluted chromatographic peak shapes and mass spectra (also not shown for brevity), suitable for quantification (results in Table 5.1). The mass spectral match factors for the intermediate-resolution PARAFAC results indicated good qualitative identification of the components of interest with similarity match values greater than 900 achieved. The biases obtained and reported in Table 5.1 indicate a decreased level of accuracy compared to the high-resolution case with the biases ranging from -5.8% to +3.6%. The %RSDs for the

intermediate-resolution case are less than or equal to 2% (Table 5.1). This indicates that although there was some redistribution of signal from the PARAFAC deconvolution, the results are still quite precise even if the accuracy is suffering due to the low multivariate selectivity achieved by some isomers.

5.4.3 Low-Resolution Butyl Benzenes Deconvolution

An extremely challenging case was then studied with the resolution between adjacent isomers on column 1 nominally 0.25 and a resolution of 0.1 on column 2 (Figure 5.3C). Only 1 second on column 1 and only 20 ms on column 2 separated the peak maxima. The deconvoluted chromatographic peak shapes, shown for one data set in Figure 5.4A and B, were quite satisfactory, although they exhibited slight deviations from the true peak shapes. The column 1 and 2 peak shapes resulting from the PARAFAC deconvolution of the high- and intermediate-resolution cases resulted in slightly better peak shapes than that shown in Figure 5.4A and B. The peak shape deviations resulting from the low-resolution case also resulted in a more significant redistribution of signal, or bias, as summarized in Table 5.1. The biases indicate a decreased level of accuracy compared to the high-resolution case with the biases ranging from -14% to +22%. The %RSD for the low-resolution case are less than 6% indicating more error in the modeling compared to the high-resolution case (Table 5.1). The deconvoluted mass spectra for the four butyl benzene isomers for the low-resolution case are shown in Figure 5.5. The spectral match factors for the PARAFAC results indicated good qualitative identification with similarity match factors of 898 or greater even with the very low chromatographic resolution. For the low-resolution case, the comparisons between NIST library spectra and the four deconvoluted spectra display a high degree of similarity (with the NIST

spectra not shown for brevity). Additionally, it can be seen by comparing iso-butyl and n-butyl benzenes (IB and BB of Figure 5.5) that the pure component spectra are highly similar and in fact only achieve a multivariate selectivity of ~ 0.12 each in the mass spectral dimension.

5.4.4 Multivariate Selectivity

Similar analyses were carried out on six additional cases for a total of nine constructed butyl benzene isomer data sets with variable resolution in both chromatographic dimensions. Approximate resolutions for column 1 were 0.25, 0.6, and 1.0. Column 2 chromatographic resolutions were nominally 0.1, 0.2, and 0.6. At each of the three nominal column 1 resolutions, the four isomers were arranged at each of the three different column 2 resolutions resulting in the nine different cases, three of which were just discussed above in detail. PARAFAC was used to analyze the six additional cases as described above. Since each of the nine cases contained four components, this resulted in 36 data points with overall multivariate selectivities ranging from 0.0054 to 0.768. The absolute value of the bias was plotted versus multivariate selectivity (Eq. (5.2)) for all four components in all nine cases (Figure 5.6A). As can be seen from Figure 5.6A, as the multivariate selectivity decreases from the maximum of ~ 0.8, the bias results are all relatively constant, below a bias of 3%, until the selectivity reaches ~ 0.12. At this point the results are more unreliable, although they still follow a general trend resulting in greater bias with less multivariate selectivity. A plot of %RSD vs. multivariate selectivity shows a similar trend (Figure 5.6B); however, the %RSD remains below 3% to a much lower level of multivariate selectivity (~ 0.005). This indicates that PARAFAC results are consistent for replicate analyses, even if the accuracy is not as optimal. As can be seen from the butyl benzene isomer case studies, TLD-initiated PARAFAC analysis can

result in highly accurate and precise results for well-resolved analytes, but it begins to suffer with extreme chromatographic overlap and low mass spectral selectivity. This emphasizes that some selectivity in each dimension needs to be present, and if there is only a very small amount of selectivity in all dimensions, the results will be less than optimal. As a practical guideline from this study, an overall multivariate selectivity of greater than 0.18 results in acceptably accurate and precise results for the deconvolution of overlapped signals. For most cases, there will be more mass spectral selectivity than that for iso-butyl and n-butyl benzenes (~ 0.12), hence PARAFAC provides a useful tool to identify and quantify constituents in a sample.

5.4.5 Metabolic Plant Extracts

To demonstrate the deconvolution technique on a real, complex sample, metabolic plant extracts were analyzed. The derivatized metabolic extracts of Huilmo (*Sisyrinchium striatum)* result in a highly complex data set as seen in the chromatogram obtained at *m/z* 73, which is a characteristic ion for the trimethylsilyl group -$(CH_3)_3Si$ associated with the derivatization products (Figure 5.7A). A relatively small portion of the entire chromatogram for selective mass channel *m/z* 73 is depicted in Figure 5.7B to highlight the complexity of the sample. There are several instances of signal overlap in the sample where a multivariate deconvolution technique could be useful. A region of the sample that was analyzed with PARAFAC is shown in Figure 5.7C. There are three overlapping components posing a challenging case for TLD-initiated PARAFAC, with very little chromatographic resolution between peaks labeled 1 and 2 on column 1 and between peaks 2 and 3 on column 2 (Figure 5.7C). A TLD model was built with 4 components to account for baseline and background offsets (results not shown for brevity). Often in complex samples, the best results are obtained by performing some type of

baseline correction on the data. This can be done in a traditional manner or by giving the model an extra component to account for baseline and background offset. The TLD results were used to initiate PARAFAC with unimodal and non-negative constraints on column 1 and 2 and only non-negative constraints in the mass spectral dimension. PARAFAC deconvolution provided successful deconvolution of the three components of interest (Figure 5.8A and B). The fourth component in the model consisting of baseline offset and noise was omitted for clarity. The deconvoluted mass spectra for the three analytes are depicted in Figure 5.8C. Deconvoluted spectrum 1 resulted in a reasonable match with the derivatized monosaccharide methyl 2,3,5,6-tetrakis-O-(trimethylsilyl)-α-D-Glucofuranoside (Match: 812, Reverse: 819, Probability: 23.6) shown in Figure 5.8D as a typical example. This and other sugars are probable components in plant metabolite extracts. The other two deconvoluted spectra did not result in as high a similarity match in the NIST mass spectral database most likely due to the obscurity of the compounds and/or the incompleteness of the database. Standards for these compounds were not currently available for this study. All three deconvoluted spectra were the most similar to isomeric derivatized monosaccharides akin to the one depicted in Figure 5.8D. Most important here is that the deconvoluted spectra are very similar (Figure 5.8C), do not contain selective ions, and are most likely isomers. Yet, PARAFAC was still able to successfully deconvolute the three compounds. Further analysis of standards and a user-compiled library is necessary for unambiguous identification and quantification of these compounds.

Based on the success of PARAFAC in this and other studies [Sinha et al. 2004a], a four-step method to identify and quantify analytes of interest is summarized here. Two data files are collected: A "sample" and a "sample +

standard addition," in which all the analytes of interest are spiked into the standard addition case. The region around each analyte of interest in the sample data set and the standard addition data set are analyzed by TLD-initiated PARAFAC, resulting in individual chromatographic peak profiles and mass spectra for both data sets. The pure chromatographic retention times and mass spectra are selective data for identification of analytes. Comparing the deconvoluted mass spectra from the sample and the standard addition cases identifies the analytes in the sample data set. This could improve the quality of similarity matches because the reference spectrum was obtained on the same instrument as the sample as opposed to the NIST library spectra that are obtained on a number of different instruments resulting in different fragmentation ion ratios. Quantification occurs by reconstructing the analyte of interest in both the sample and the standard addition and applying the usual techniques for quantification via standard addition. The problems associated with small retention time shifts between sample and standard separations are eliminated because the deconvolution and identification are performed separately on both the sample and the standard addition, thus simplifying the method substantially.

5.5 Conclusions

It was demonstrated that the GC x GC-TOFMS trilinear data structure is compatible with third-order chemometric analysis techniques such as TLD and PARAFAC. This was true for both valve modulated and thermally modulated GC x GC instruments. The effects of altering chromatographic resolution on the results of a PARAFAC analysis was investigated with constructed data sets of four isomers that exhibit similar mass spectra. Two of the compounds had extremely similar mass spectra with mass spectral multivariate selectivities of only ~ 0.12. It was shown that PARAFAC leads to

successful qualitative identification of isomeric analytes with a wide range of selectivities (0.0054 to 0.768), but quantitative results were better for cases where the overall multivariate selectivity for a given analyte was greater than 0.18. The ability of PARAFAC to successfully deconvolute isomers was demonstrated on three overlapping species of possibly isomeric monosaccharide derivatives from a complex plant metabolite sample of Huilmo (*Sisyrinchium striatum*). Quantification of analytes of interest, including isomers, can therefore be identified and quantified using PARAFAC without selective ions, peak shape predictions or retention time alignment between the sample and the standard prior to analysis. PARAFAC is easily automated and could potentially be applied to complete GC x GC-TOFMS chromatograms in the way NIST's Automated Mass Spectral Deconvolution and Identification System (AMDIS) [Halket et al. 1999; Stein 1999] is applied to GC-MS chromatograms. AMDIS takes a GC-MS chromatogram, identifies the location of all peaks, deconvolutes the unresolved peaks and then searches a library for matching compounds. A version of AMDIS extended to GC x GC-TOFMS and PARAFAC could work as complementary peak deconvolution techniques. While AMDIS relies on selective ions, PARAFAC relies on some selectivity in each dimension.

Table 5.1. Accuracy and precision studies for TLD-initiated PARAFAC deconvolution. Results for the high-, intermediate-, and low-resolution constructed data sets comprised of iso-butyl, sec-butyl, tert-butyl, and n-butyl benzene isomers for 3 replicates. "H," "I," and "L" refer to the high-resolution, the intermediate-resolution, and the low-resolution cases respectively. The selectivity corresponds to the overall multivariate selectivity, which is the product of the selectivities in each dimension (Eq. (5.2)).

Analyte	**Selectivity Eq. (5.2)**			**%Bias**			**%RSD**		
	H	**I**	**L**	**H**	**I**	**L**	**H**	**I**	**L**
Sec-butyl Benzene	0.768 ± 0.009	0.366 ± 0.004	0.05 ± 0.01	+1.5	+0.9	+0.9	1.1	1.0	1.3
Tert-butyl Benzene	0.659 ± 0.003	0.29 ± 0.02	0.044 ± 0.001	+0.6	+1.3	-8.0	0.8	0.8	2.4
Iso-Butyl Benzene	0.1203 ± 0.0007	0.058 ± 0.001	0.0056 ± 0.0005	+0.6	+3.6	-14	0.8	1.4	6.4
n-Butyl Benzene	0.114 ± 0.002	0.059 ± 0.001	0.0054 ± 0.0005	-3.0	-5.8	+22	0.8	2.0	4.0

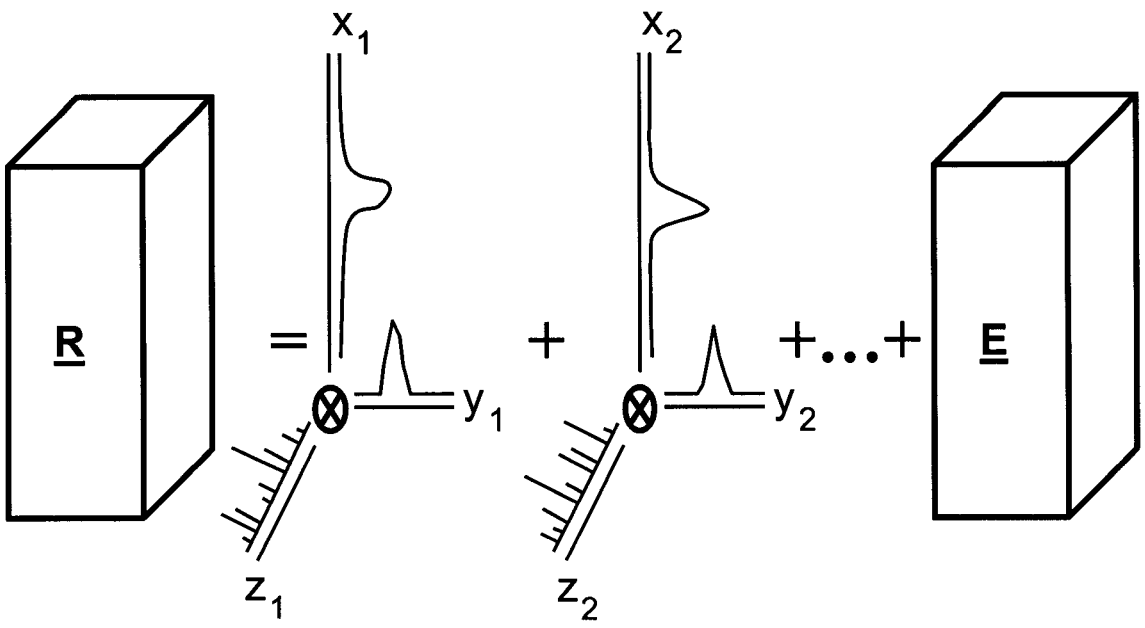


Figure 5.1. Illustration of the trilinear data structure of GC x GC-TOFMS data. For the instrument response $\underline{\mathbf{R}}$ there are distinctive profiles in both chromatographic dimensions ($\mathbf{x}_n$ and $\mathbf{y}_n$) and a unique spectrum ($\mathbf{z}_n$) for each constituent in a data matrix. The error (e.g., noise) is denoted $\underline{\mathbf{E}}$. This is mathematically described as $\underline{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}_n \otimes \mathbf{y}_n \otimes \mathbf{z}_n + \underline{\mathbf{E}}$.

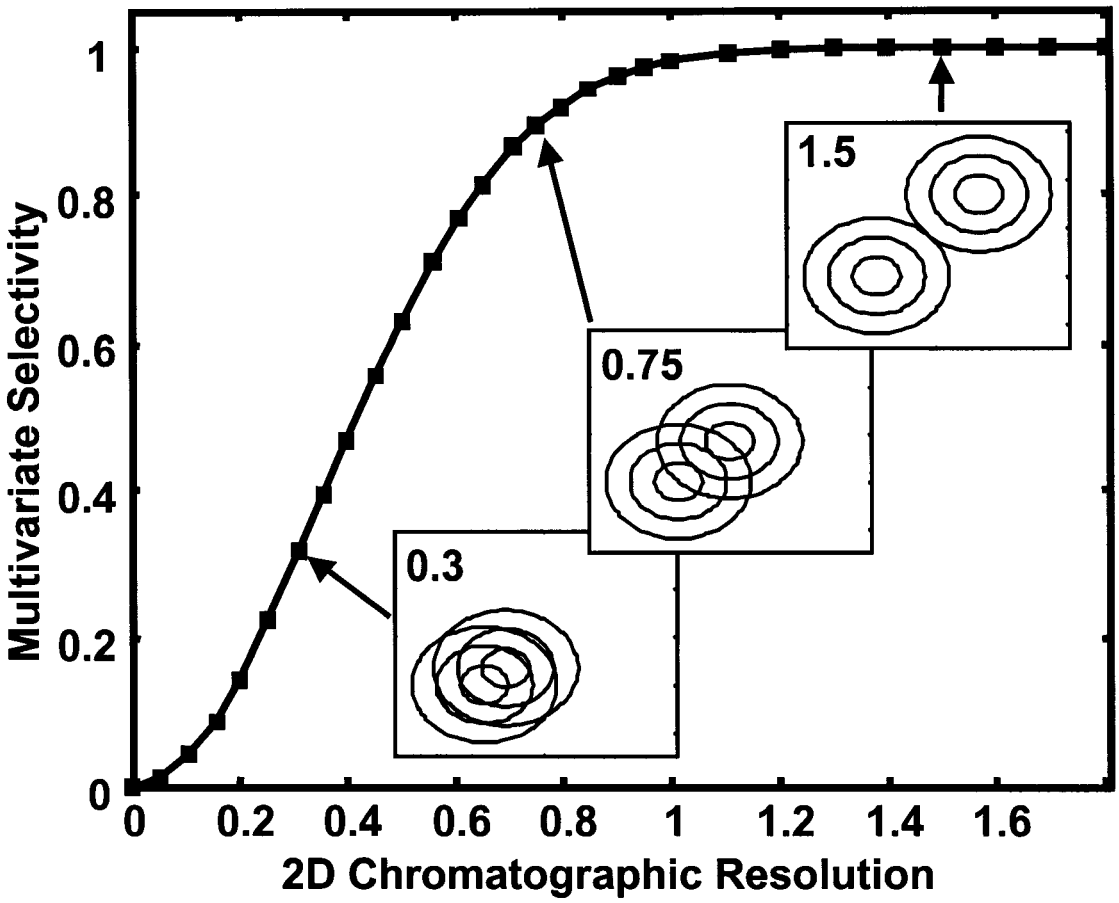


Figure 5.2. Plot of data multivariate selectivity as a function of 2D chromatographic resolution.

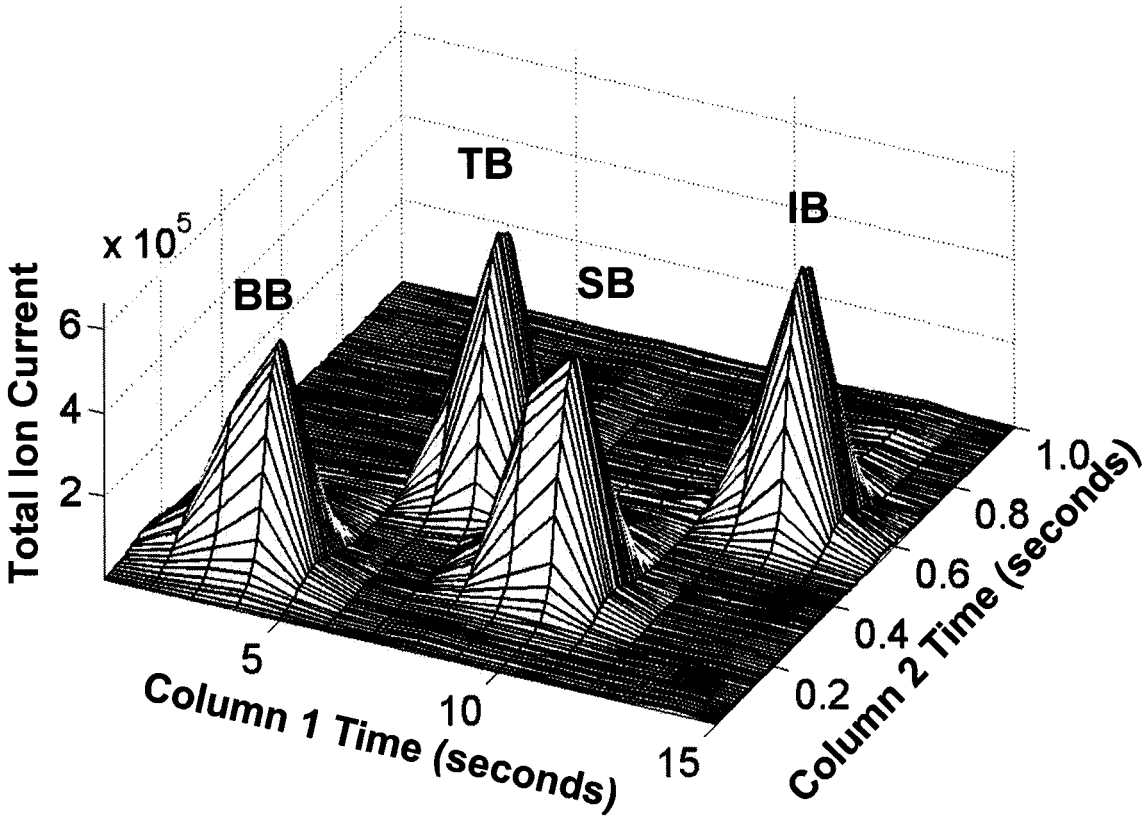


Figure 5.3A. 3D TIC image of the high-resolution constructed data set. Isomeric compounds in the data set are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

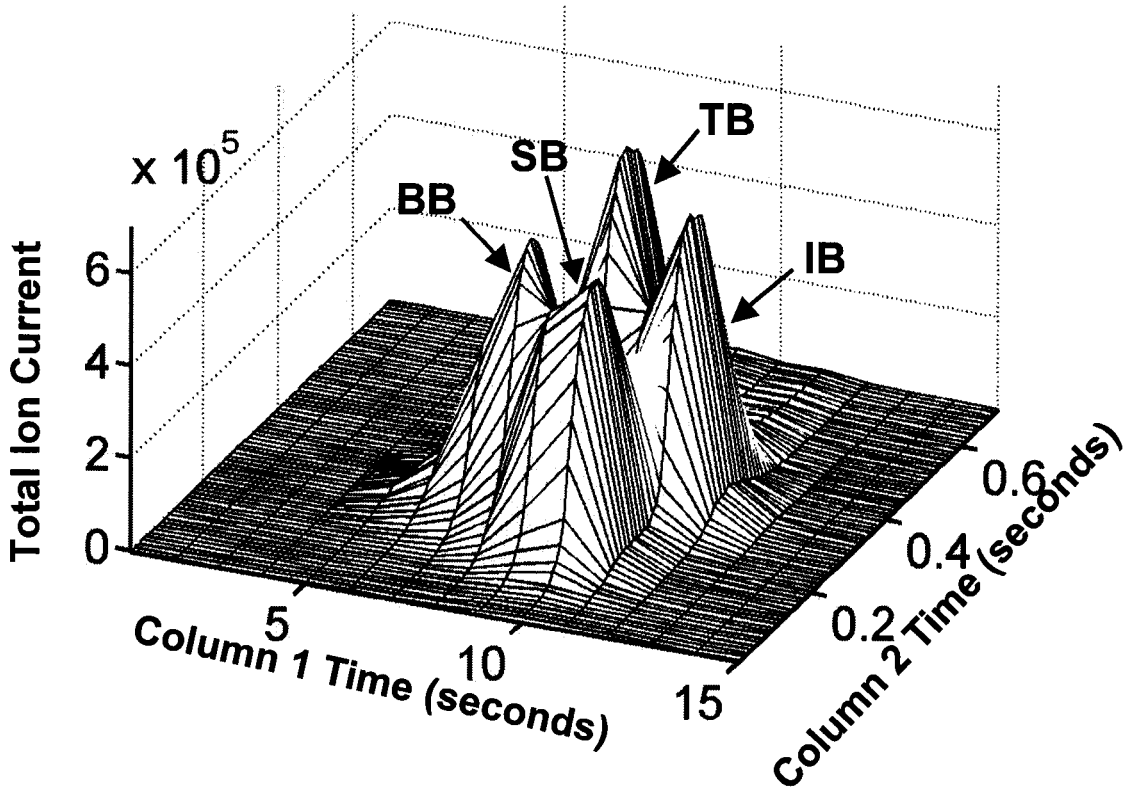


Figure 5.3B. 3D TIC image of the intermediate-resolution constructed data set. Isomeric compounds in the data set are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

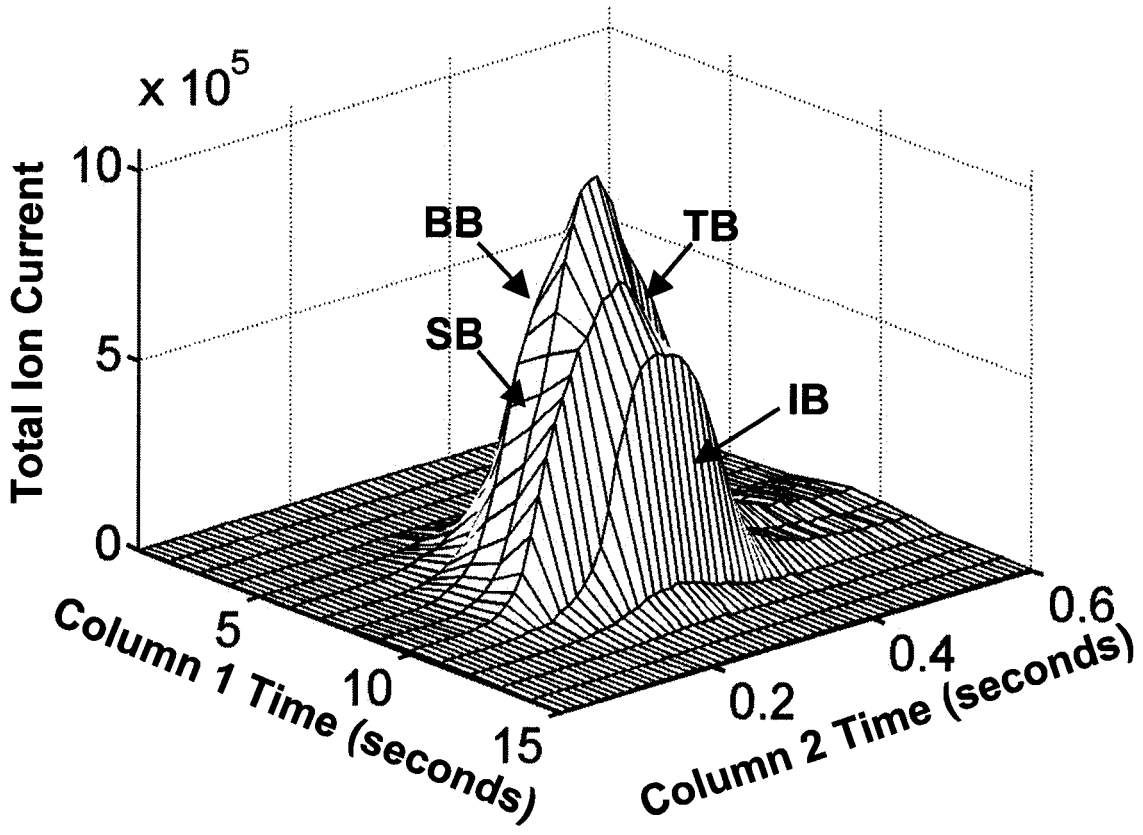


Figure 5.3C. 3D TIC image of the low-resolution constructed data set. Isomeric compounds in the data set are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

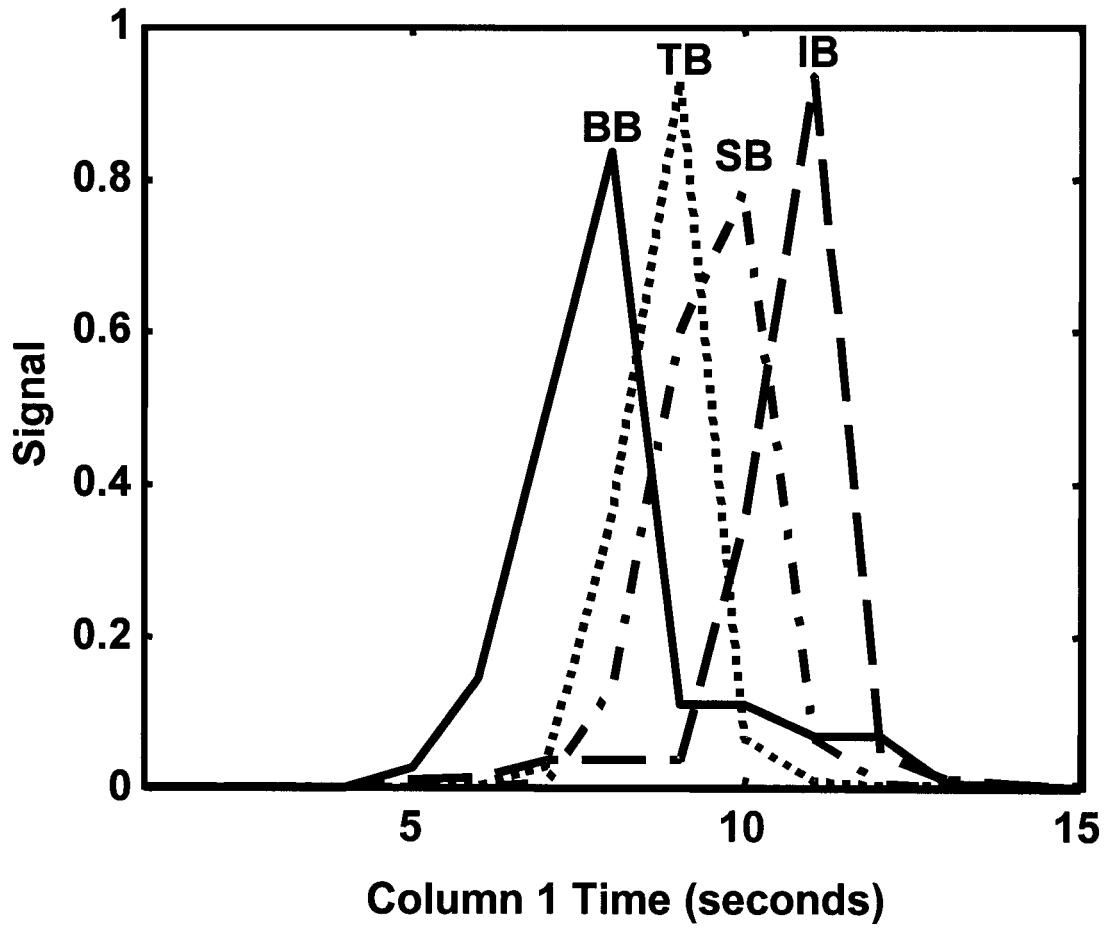


Figure 5.4A. PARAFAC deconvoluted column 1 profiles for the low-resolution isomers. Isomeric compounds are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

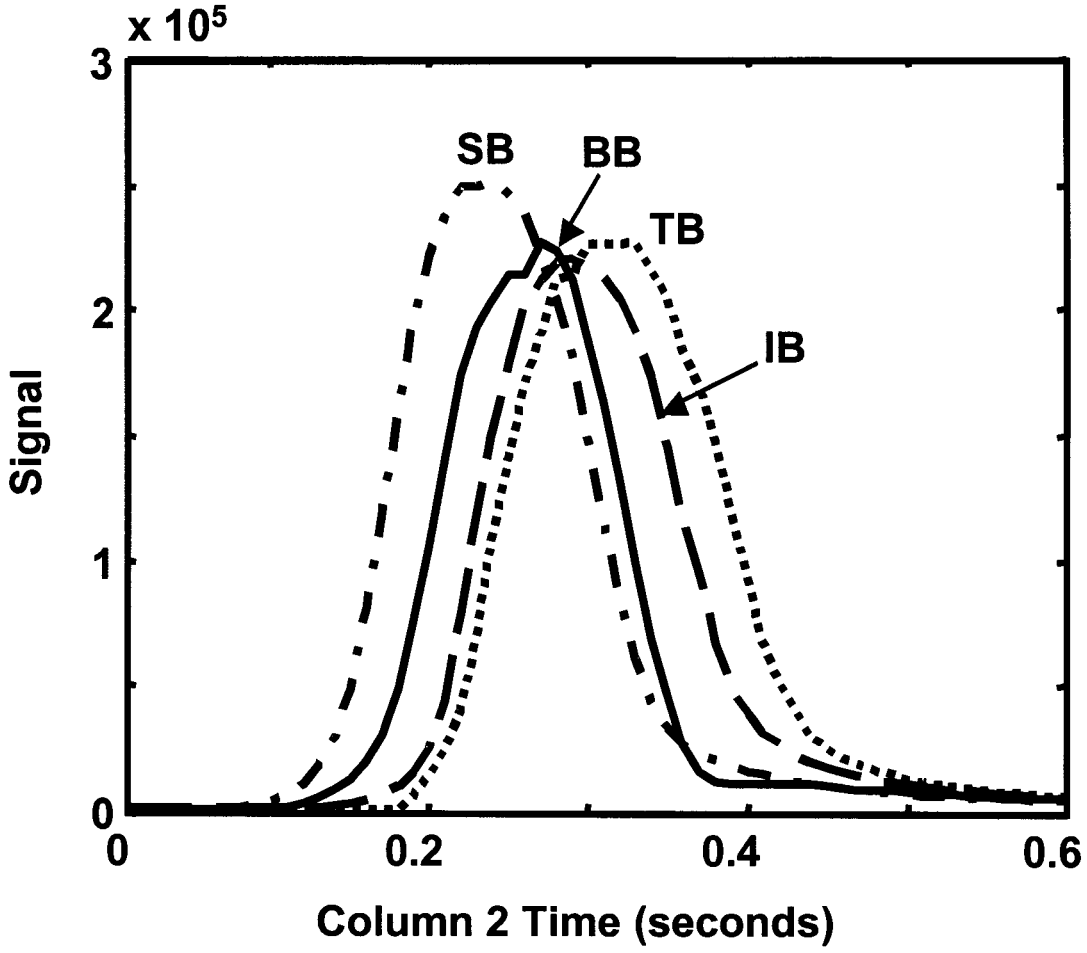


Figure 5.4B. PARAFAC deconvoluted column 2 profiles for the low-resolution isomers. Isomeric compounds are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

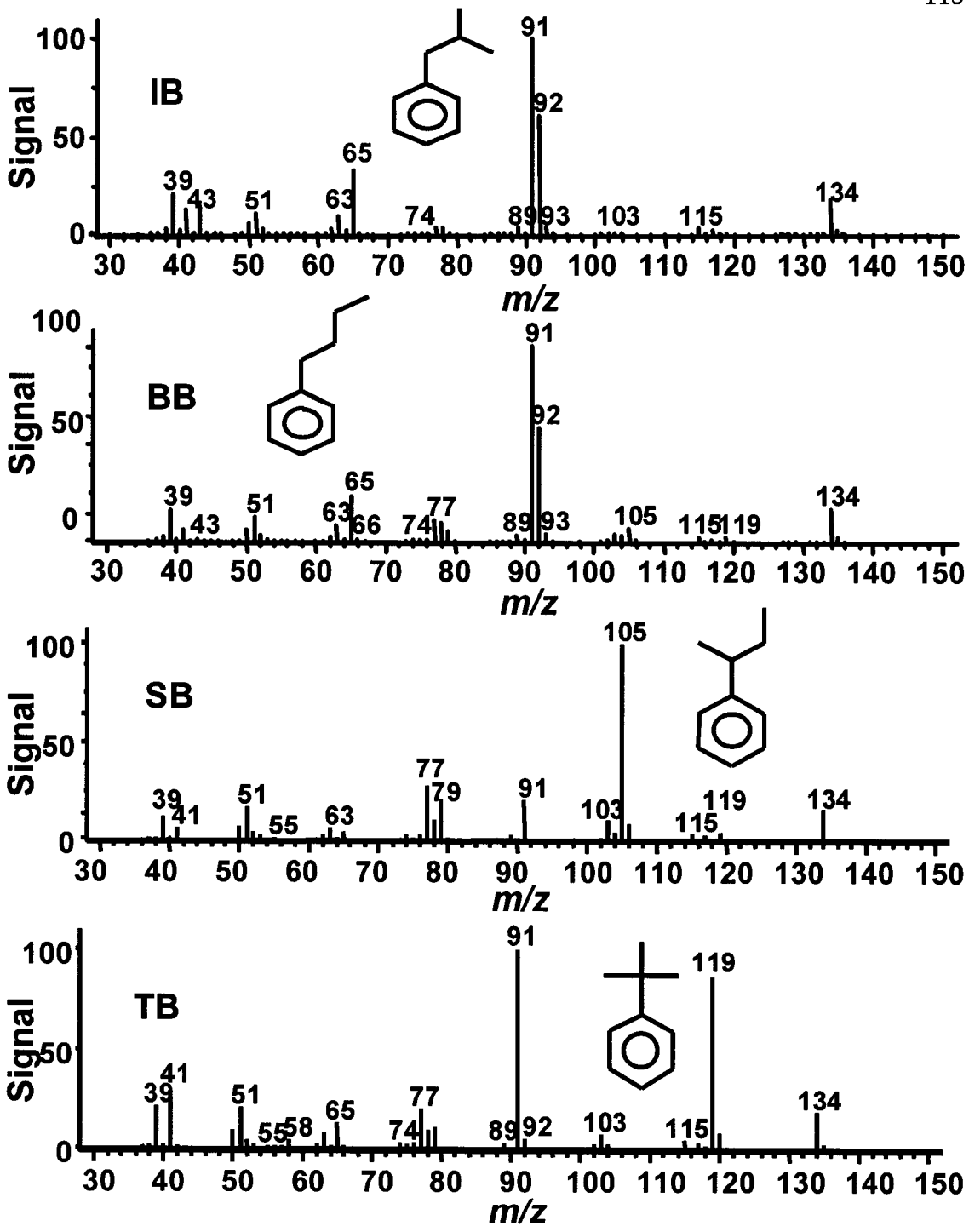


Figure 5.5. PARAFAC deconvoluted mass spectra for the low-resolution isomers. Isomeric compounds are iso-butyl (**IB**), sec-butyl (**SB**), tert-butyl (**TB**), and n-butyl (**BB**) benzenes.

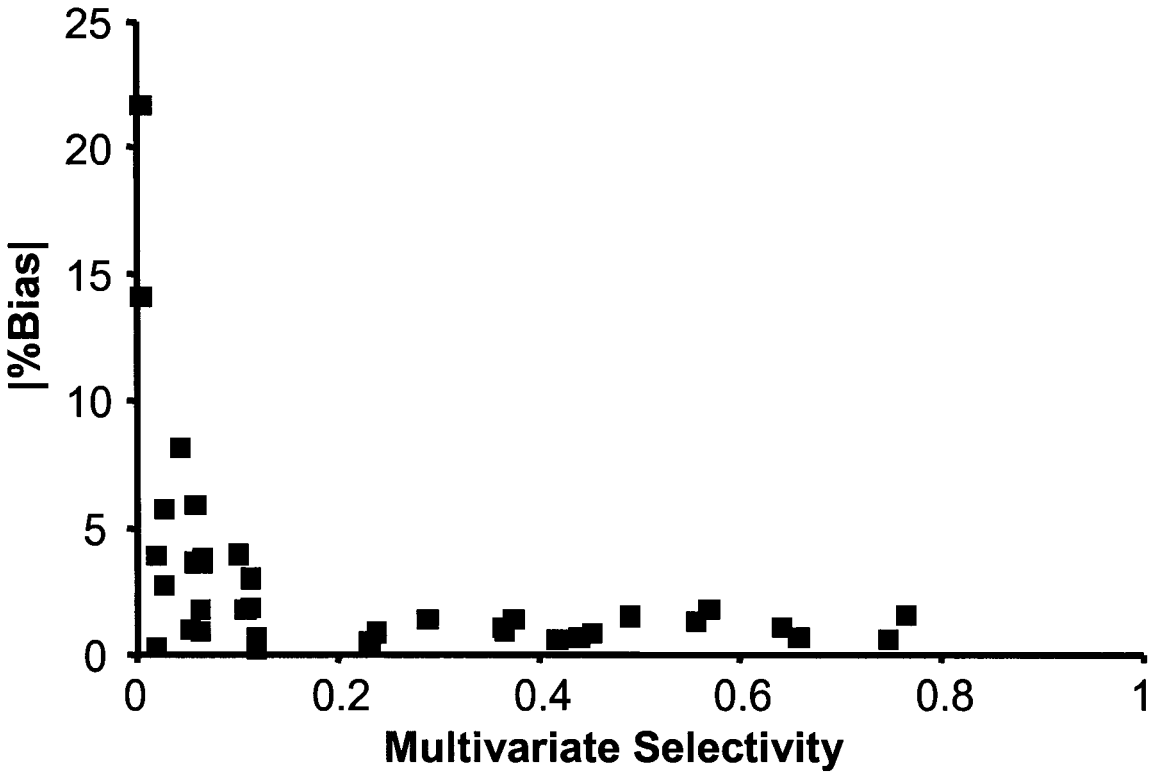


Figure 5.6A. Plot of the absolute value of the %Bias of PARAFAC deconvolution results. Data points are for the four compounds (iso-butyl, sec-butyl, tert-butyl, and n-butyl benzenes) in nine different chromatographic resolution cases versus multivariate selectivity. Three replicates were obtained for each compound in each of the nine cases. The average standard deviation of the multivariate selectivities for all compounds and cases was ± 0.007.

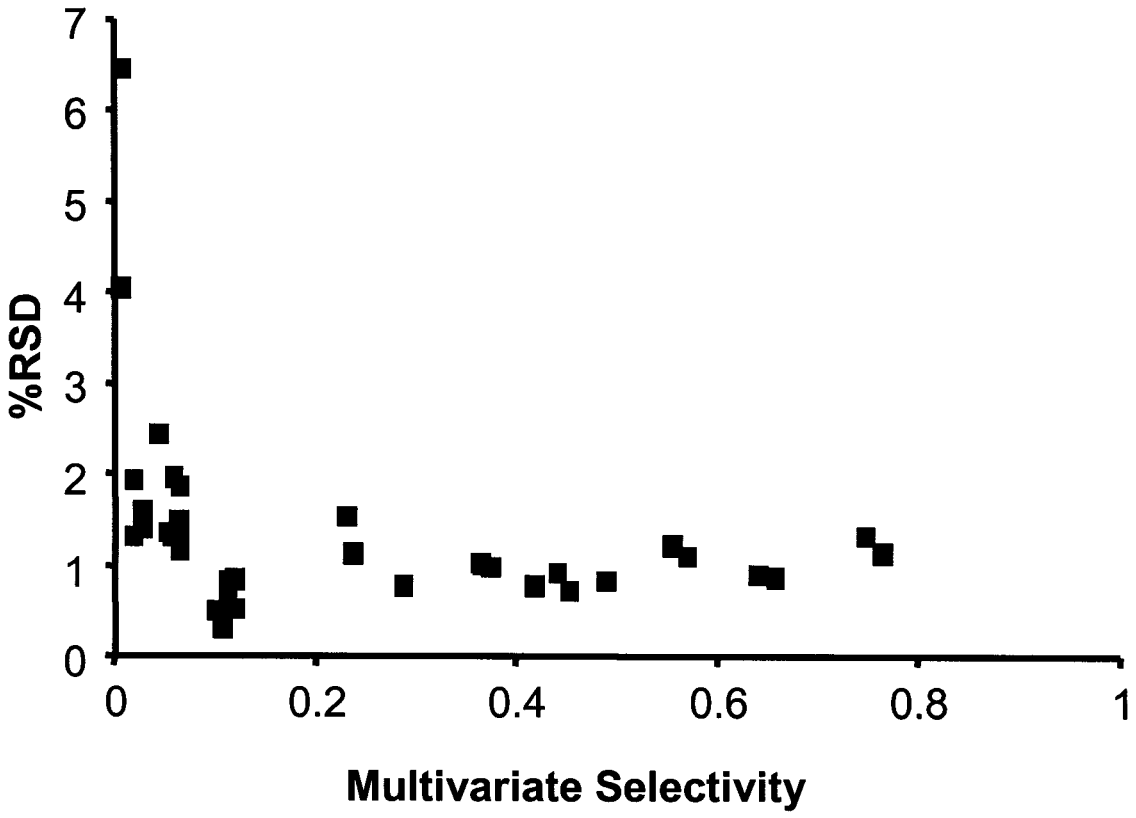


Figure 5.6B. Plot of the %RSD of PARAFAC deconvolution results. Data points are for the four compounds (iso-butyl, sec-butyl, tert-butyl, and n-butyl benzenes) in nine different chromatographic resolution cases versus multivariate selectivity. Three replicates were obtained for each compound in each of the nine cases. The average standard deviation of the multivariate selectivities for all compounds and cases was ± 0.007.

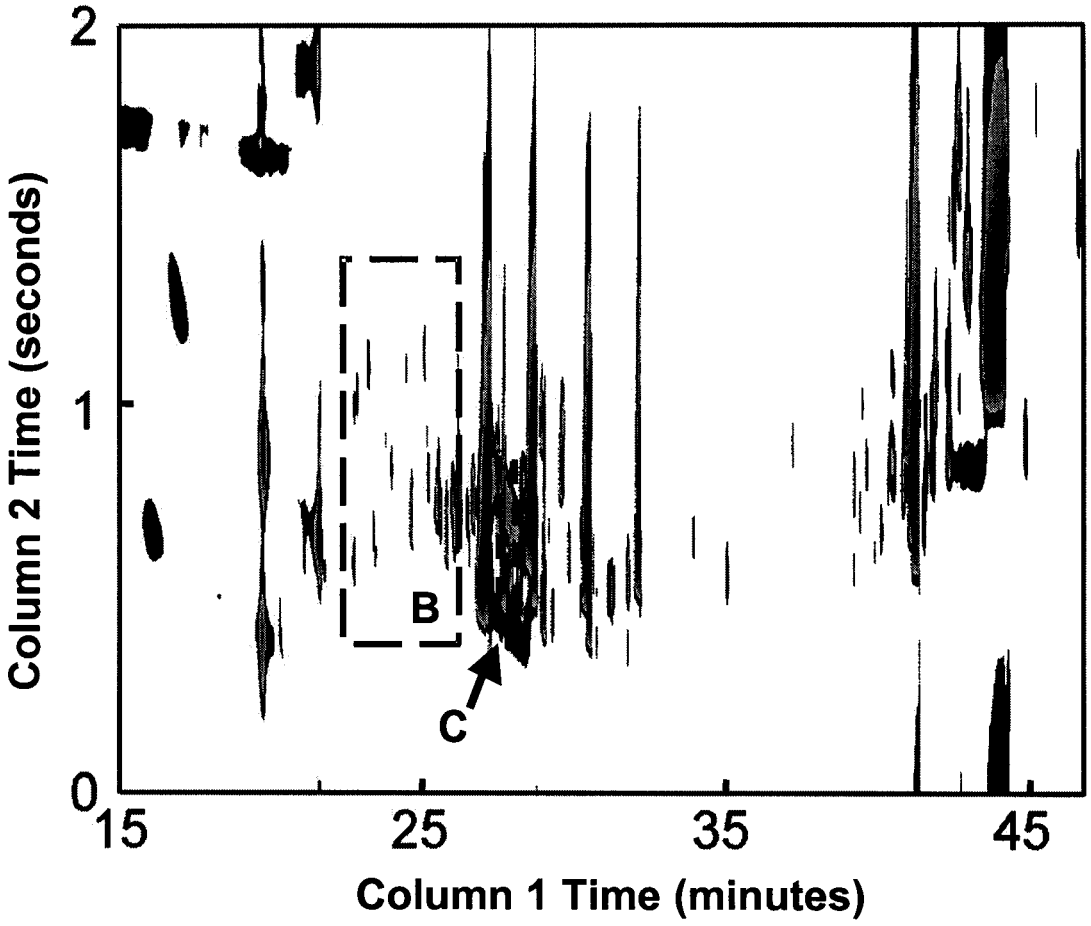


Figure 5.7A. GC x GC-TOFMS analysis of the TMS derivatized Huilmo (*Sisyrinchium striatum*) plant extract. Chromatogram of *m/z* 73 is shown. Region appearing in Figure 5.7B is denoted by dashed box with label '**B**'. Region used for PARAFAC analysis depicted in Figure 5.7C is denoted by dashed box with label '**C**'.

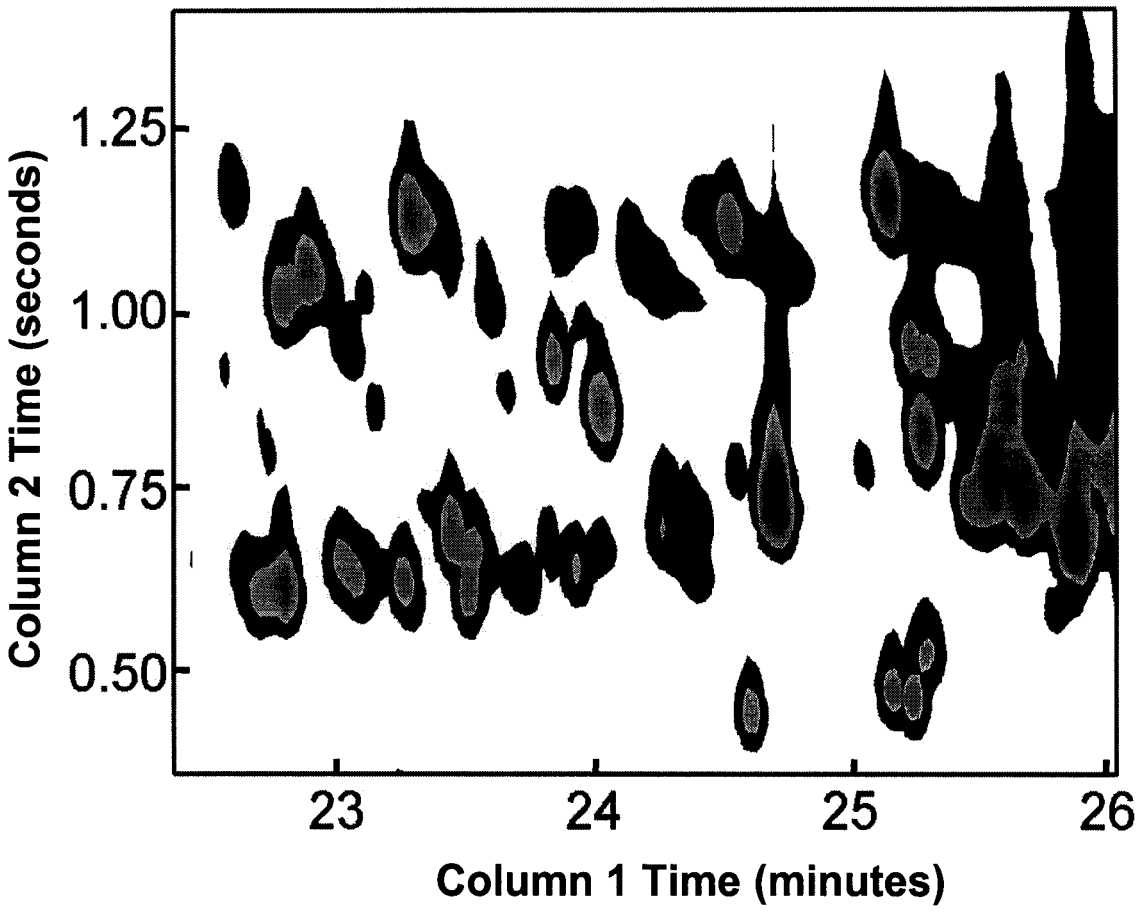


Figure 5.7B. Complex sub-region of the TMS derivatized Huilmo extract. Chromatogram of *m/z* 73 is shown.

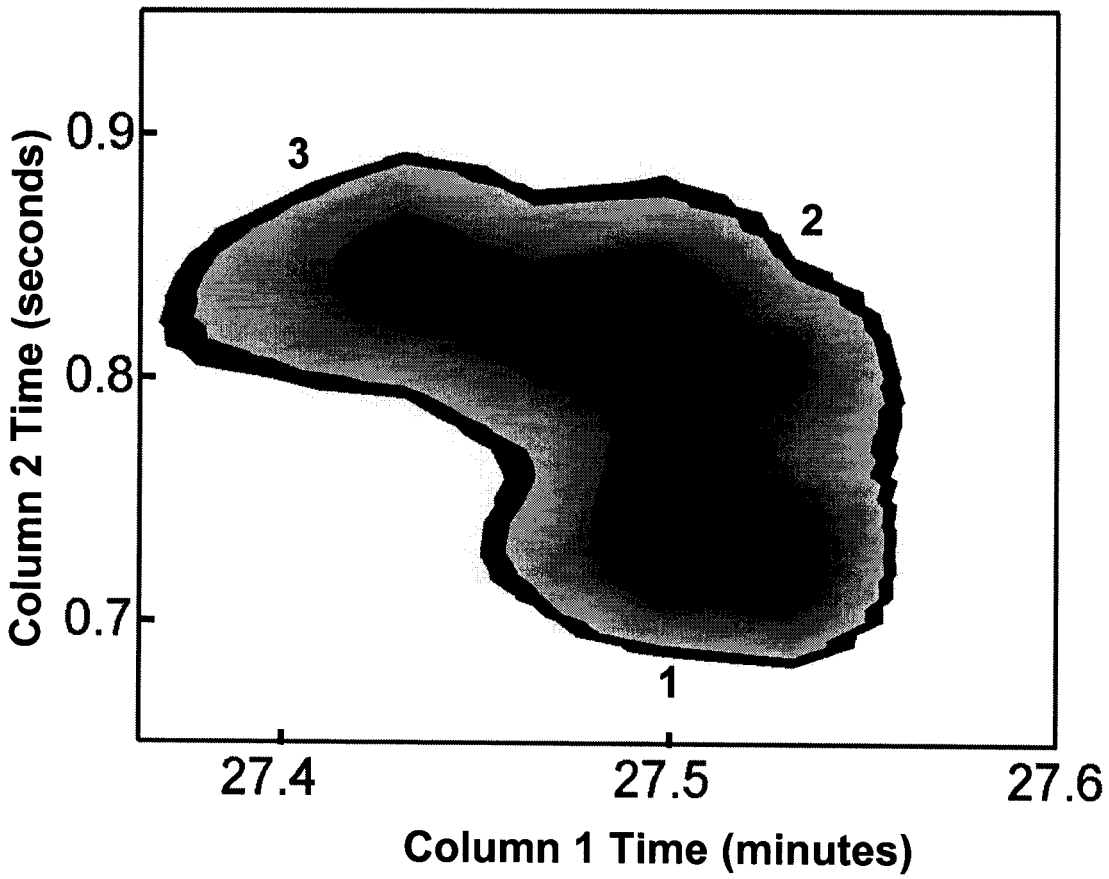


Figure 5.7C. Region of the TMS derivatized Huilmo extract analyzed by PARAFAC. The TIC chromatogram is shown. The three overlapping signals were analyzed using a 4-component model.

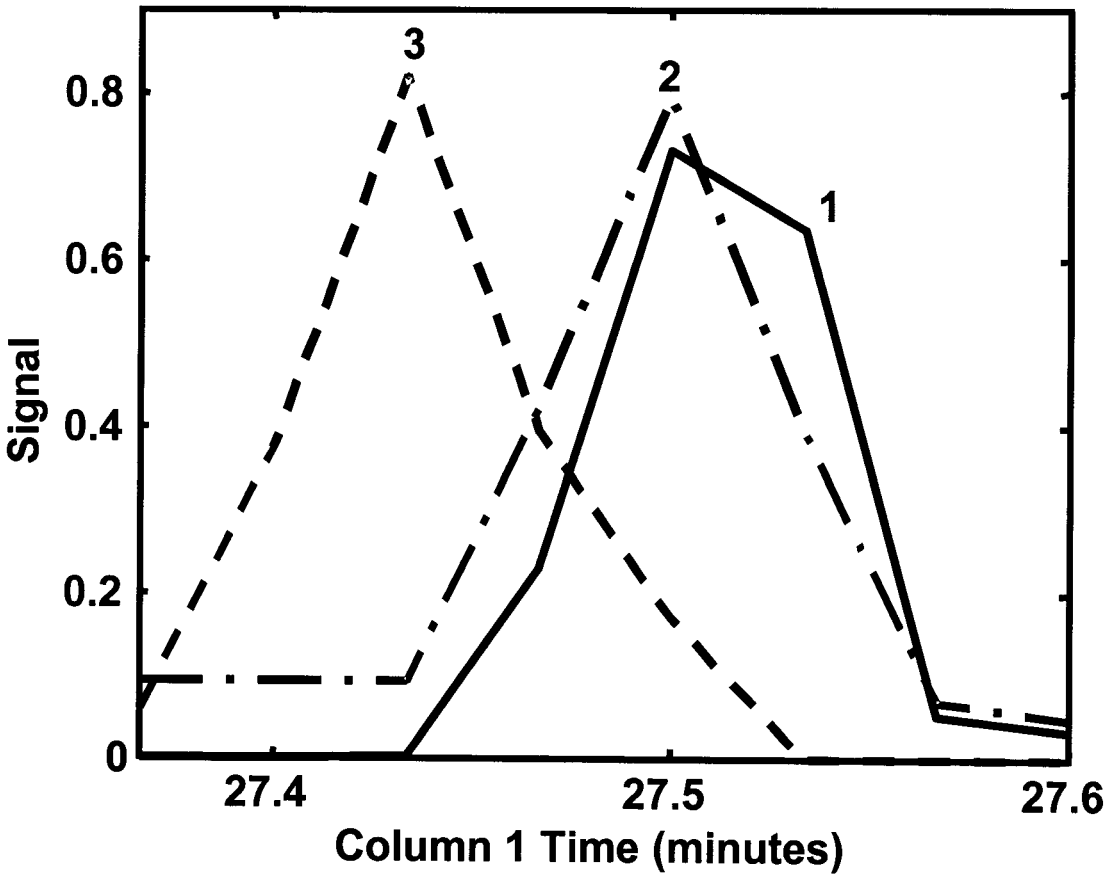


Figure 5.8A. PARAFAC deconvoluted column 1 pure component profiles. Three species labeled **1**, **2**, and **3** shown (component 4 consisting of baseline offset omitted for clarity).

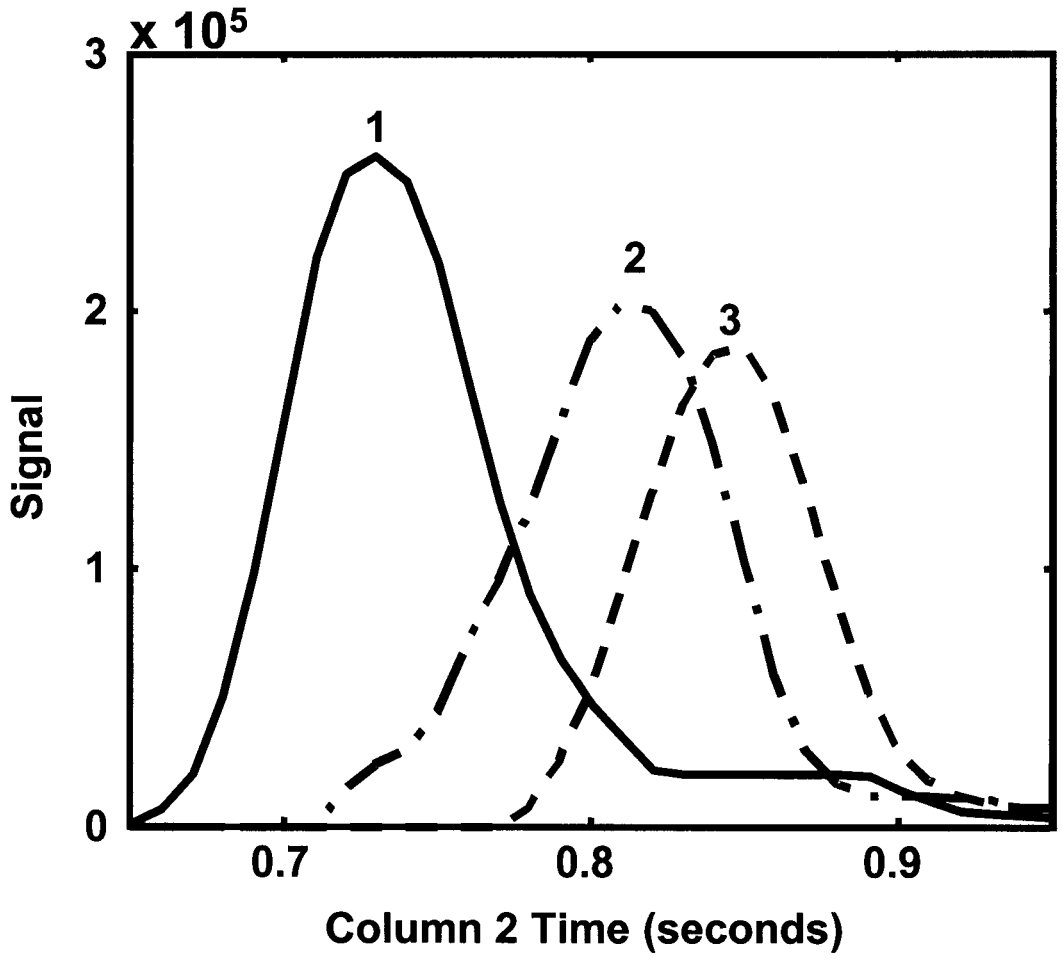


Figure 5.8B. PARAFAC deconvoluted column 2 pure component profiles. Three species labeled **1**, **2**, and **3** shown (component 4 consisting of baseline offset omitted for clarity).

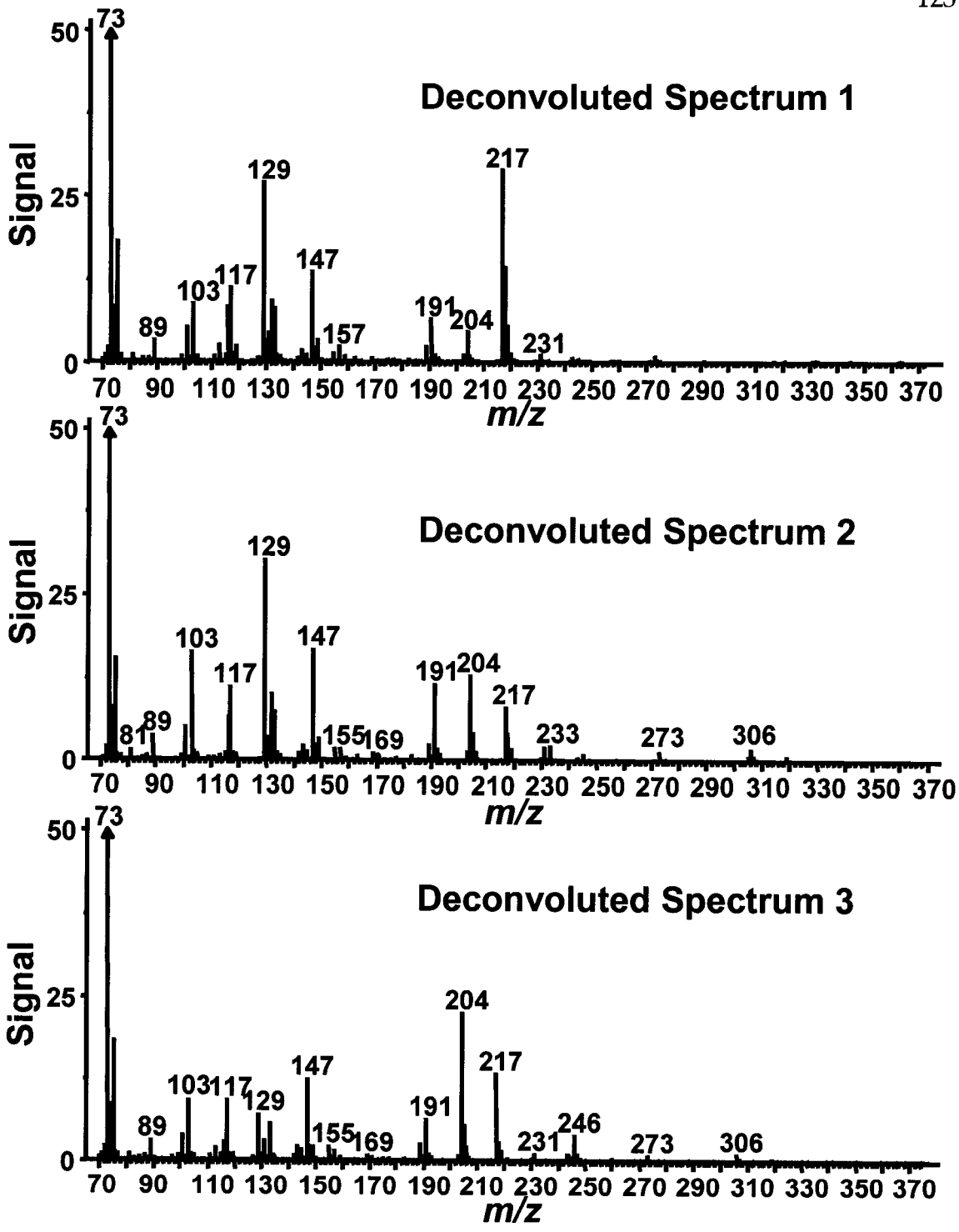


Figure 5.8C. PARAFAC deconvoluted mass spectral profiles.

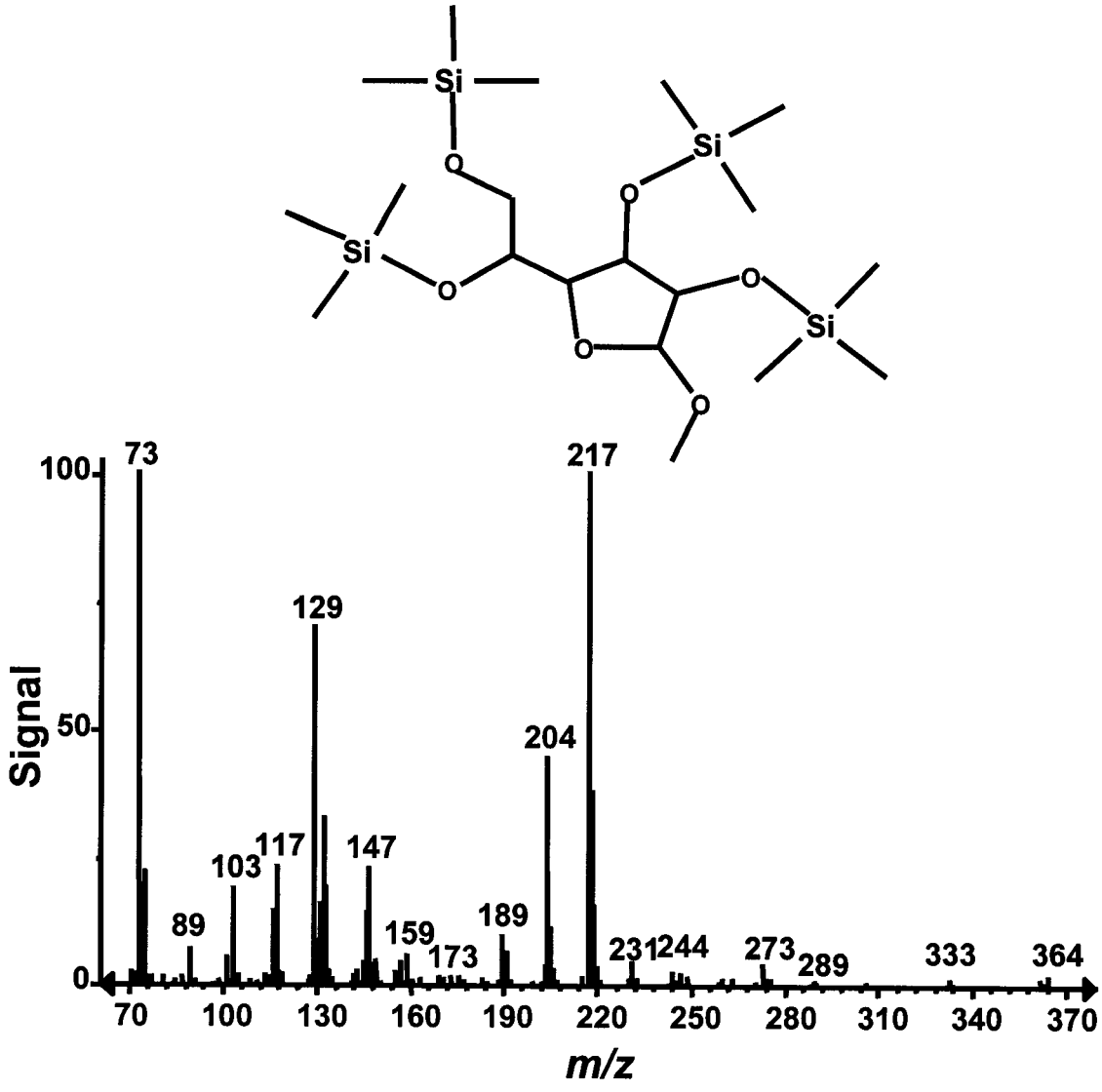


Figure 5.8D. Best NIST library mass spectrum similarity match for deconvoluted species **1**. Compound: methyl 2,3,5,6-tetrakis-O-(trimethylsilyl)-α-D-Glucofuranoside (match: 812, reverse: 819, probability: 23.6).

Chapter 6. Comparison of Deconvolution Methods Applicable to GC x GC - TOFMS Data

6.1 Introduction

Research and development of the advantageous combination of GC x GC with TOFMS has been expanding in recent years [Adahchour et al. 2003; Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; de Koning et al. 2003; Dimandja 2003; Focant et al. 2003; Korytar et al. 2003; Lu et al. 2003; Shellie et al. 2001b; Sinha et al. 2003b; Sinha et al. 2004b; van Deursen et al. 2000; Welthagen et al. 2003; Zrostlikova et al. 2003]. GC x GC-TOFMS is an excellent example of a state-of-the-art hyphenated instrument that creates a nearly unmanageable amount of data, yet has the potential to more completely describe complex chemical mixtures. One potential method for extracting more useful data from these complex data sets is through peak deconvolution.

Most peak deconvolution methods used for GC x GC-TOFMS data analysis are typically extensions of those that are already used for GC-MS and do not utilize the trilinear data structure. Two techniques of this type are the Automated Mass Spectral Deconvolution and Identification System (AMDIS) and the LECO automated peak finding and deconvolution ChromaTOF software. These deconvolution methods extend GC x GC-TOFMS data into a series of column 2 separation dimension GC-MS data. Each column 2 chromatogram is analyzed independently. The LECO ChromaTOF software then recombines the analyzed data along the column 1 separation dimension. Both of these methods rely on selective ions to generate peak model shapes and essentially do not utilize the column 1 separation dimension during deconvolution of overlapped peaks.

Trilinear chemometric techniques have been developed that specifically utilize the data structure found in GC x GC-TOFMS, which is not found in GC-MS data. Third-order chemometric techniques such as trilinear decomposition (TLD) and parallel factor analysis (PARAFAC) have been successfully applied to GC x GC-TOFMS data to deconvolute overlapped signals because of the inherent trilinear data structure [Sinha et al. 2004a; Sinha et al. 2004b]. With PARAFAC, the information along all three dimensions is analyzed simultaneously, although some selectivity is needed in each of the three dimensions for peak deconvolution using a single data set.

There has been no evaluation, however, of the benefits and drawbacks of available deconvolution methods for GC x GC-TOFMS data. In order to compare the deconvolution results generated by PARAFAC, AMDIS, and the LECO ChromaTOF software, an environmental sample containing fuel components, pesticides and natural products was separated with GC x GC-TOFMS. With the goal of providing a straightforward, direct comparison of the deconvolution techniques, not a comprehensive evaluation, one representative region of overlapping peaks in the sample was analyzed in triplicate with the three deconvolution methods. A proprietary test compound was spiked into the sample matrix prior to analysis. This test compound provided the benchmark for the deconvolution comparison. It will be demonstrated that all three methods are ultimately able to deconvolute a comparable, identifiable spectrum for the test compound from initially partially resolved data. The specific results for each method, however, differ in the ease-of-use, chromatographic profiles generated, time necessary for deconvolution, and quantitative potential.

6.2 Background

6.2.1 TLD and PARAFAC

TLD and PARAFAC are chemometric techniques that have been well characterized in the literature [Booksh et al. 1994; Bro 1997; Faber et al. 2003; Li and Gemperline 1993; Sánchez and Kowalski 1990]. TLD is an eigenvalue-based solution to the trilinear PARAFAC model [Faber 2002; Sánchez and Kowalski 1990]. The alternating least squares (ALS) based solution to the trilinear PARAFAC model has acquired the name PARAFAC and is the most popular method. PARAFAC deconvolution employs a starter solution followed by ALS to find a solution to the model [Bro 1997; Faber et al. 2003]. Possible starter solutions for the least squares method include TLD, singular values, random values, and random orthogonalized values. TLD initialization provided the fastest approach and gave the best results for GC x GC-TOFMS data [Sinha et al. 2004a; Sinha et al. 2004b]. TLD is advantageous because it does not require a starter solution and because it is computationally fast; however, TLD followed by PARAFAC deconvolution has been demonstrated to give better results than TLD alone [Sinha et al. 2004a; Sinha et al. 2004b]. The strength of PARAFAC deconvolution is mainly attributed to non-negative and unimodal constraints that are incorporated into the PARAFAC deconvolution algorithm.

PARAFAC is only applied to a small, 3D submatrix containing the analyte(s) of interest. The dimensions of this 3D matrix are often only twice the peak widths on column 1, by twice the peak widths on column 2, by all relevant mass channels. Generation of a TLD model for PARAFAC initiation requires entering the submatrix of data and the number of components present into the TLD algorithm. Baseline correction prior to analysis is not

required, but often leads to better PARAFAC results. This is partly due to the unimodal constraint, which is not a realistic constraint for baseline. Deconvoluting data without baseline correction requires the number of components entered into the model to be augmented by one. For example, if there were three visible peaks in a submatrix without baseline correction, the number of components entered into the model should be four. After TLD analysis, the submatrix, the number of components present, and the TLD results are used to generate a PARAFAC model of the data. Unimodal and non-negative constraints are optional, but often lead to more realistic and accurate results. When analyzing data without baseline correction, the unimodal constraint is sometimes not applicable, leading to an error in the model. The ultimate results of the PARAFAC model are three sets of vectors that describe the column 1 pure component profiles, column 2 pure component profiles, and pure mass spectra for each of the components in the GC x GC-TOFMS submatrix (see Figure 1.2). Concentration information for subsequent quantitation appears with the column 2 profiles and the results are normalized in the other two dimensions.

6.2.2 AMDIS

AMDIS was developed by the National Institute of Standards and Technology (NIST) to extract "pure" component spectra from complex GC-MS chromatograms [Stein 1999]. The deconvolution is based on the model peak method developed by Dromey et al. [Dromey et al. 1976]. Although AMDIS was developed for the automated identification of chemical weapons and similar compounds [Stein 1999], it has since been applied to GC-MS and LC/MS data containing compounds of many types including urinary organic acids involved in metabolic disorders [Halket et al. 1999]. AMDIS has four main analysis steps: noise analysis, component perception, spectrum

deconvolution, and compound identification. AMDIS first analyzes the background and calculates a noise level for later processing. It then analyzes the data for a maximum in a unique ion trace. If there are maxima for other ion traces at the same time, it is assumed that there is a peak at that location. A model peak shape is then formed based on the unique ion traces. In the next step it calculates a "clean" spectrum for each peak using the model peak shape and by explicitly subtracting up to two adjacent components when needed. The deconvoluted spectrum is then subjected to a library search to identify the compound. Although AMDIS is not currently configured to specifically deconvolute GC x GC-TOFMS data, the GC x GC-TOFMS data in the form of one long GC-MS chromatogram can be analyzed. To configure the data in this form, one can string each column 2 (GC-MS) chromatogram end-to-end. This data structure results in multiple peaks per compound due to column 1 effluent modulation. The results will entail a set of deconvoluted spectra for each compound corresponding to the column 2 separations resulting from the modulation frequency. For example, a compound with a column 1 peak width of 8 seconds and a GC x GC modulation frequency of 2 seconds will result in 4 modulations and subsequent separations on column 2. The AMDIS deconvolution results for this example would be a set of 4 deconvoluted spectra corresponding to the 4 column 2 modulations and separations.

6.2.3 LECO ChromaTOF Software

The LECO automated peak finding and deconvolution ChromaTOF software for GC x GC-TOFMS data is an extension of the LECO GC-TOFMS deconvolution software version in which the data are treated in a similar fashion to AMDIS. The deconvolution is based on finding a selective ion in the peak of interest and in any overlapping peaks from which model peak

shapes are used to extract pure component spectra. Where AMDIS allows up to two interfering components, the ChromaTOF software allows more than two interfering components. The automated data analysis process with the ChromaTOF software includes baseline subtraction, peak finding, peak deconvolution, and quantitative calculations for peak area and height. For accurate deconvolution, the software requires two to three full scan spectra between the apexes of adjacent peaks [Dallüge et al. 2002a; Veriotti and Sacks 2003], a selective ion in each peak with sufficient signal-to-noise [Veriotti and Sacks 2003], accurate column 2 peak width estimates, and optimally 20-30 points (i.e., full scan spectra) across a column 2 peak trace. For GC x GC-TOFMS data, once all peaks are found and deconvoluted in the raw GC-TOFMS data, the software groups the multiple deconvoluted column 2 peaks based on an accurate estimate of the column 1 peak width and the similarity of the deconvoluted spectra. The time needed to analyze an entire GC x GC-TOFMS chromatogram with this software depends on sample complexity. One analysis of cigarette smoke with a 60-minute chromatogram with a DAQ rate of 100 spectra/second resulted in a 7 hour processing time [Dallüge et al. 2002b]. For more specific target analysis, the processing times of smaller regions of a chromatogram are substantially shorter. The peak finding and deconvolution software is integrated with the software that controls the GC x GC-TOFMS so that exporting data is unnecessary.

6.3 Experimental

6.3.1 Instrumental Parameters

The analysis of environmental sample #2 was performed using a LECO Pegasus 4D GC x GC-TOFMS instrument (LECO Corporation, St. Joseph, MI, USA). Three data sets were collected to provide replicate analysis. A

proprietary test compound was spiked into the sample at a level of 262 ng/ml. Column 1 was a 60-m 250-μm i.d. capillary column with a 0.5-μm 5% diphenyl/95% dimethyl polysiloxane film (DB-5; J&W Scientific/Agilent Technologies). Column 2 was a 2-m 180-μm i.d. capillary column with a 0.1-μm 90% biscyanopropyl/10% phenylcyanopropyl film (RTX-2330; Restek Corp., Bellefonte, PA, USA). These columns were joined using a Vu2 union (Restek Corp.). Modulation was performed using cryogenic modulation. The modulator at the head of column 2 concentrated the column 1 effluent during each column 2 separation. A 0.4 second "hot pulse" made an injection onto column 2 and began each new column 2 separation when the cryogenic gas was shut off and the heated air jets (40 °C above the oven temperature) were switched on. The modulation frequency was 1.5 seconds. Ultra high purity helium (1.0 ml/min) was used as the carrier gas. 2.0 μl of environmental sample #2 was injected splitless for 0.5 min. The column 1 oven ramp began at 40°C with a hold time of 0.5 minutes, increased at 20 °C/min to 100 °C, increased at 5 °C/min to 210 °C, then increased at 10 °C/min to 260 °C, and held for 9.5 minutes. Column 2 was housed in a separate oven and ramped at the same rates to temperatures consistently 15 °C higher than the column 1 oven temperature for all but the final temperature ramp and hold. During the last ramp and hold, the column 2 oven ramped from 225°C to 260 °C at 10 °C/min and held for 11 min so as not to exceed the temperature limits of the column 2 stationary phase. No mass spectra were collected during the solvent delay for the first 6.75 minutes of each run. The transfer line was maintained at 260°C and the ion source set point was 200°C. The detector voltage was -1800 V and the filament bias was -70 V. Mass spectra were collected from *m/z* 35 - 300 at 100 spectra/second. Mass spectral similarity searches were performed using the NIST MS Search 2.0 (NIST/EPA/NIH Mass Spectral

Library; NIST 02) and a MS library built in-house that included the spectrum of the test compound for all subsequent analyses.

6.3.2 PARAFAC Analysis

Three replicate data sets for environmental sample #2 were exported as comma separated value (.csv) files and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing. The region surrounding the test compound was isolated as a submatrix for deconvolution analysis. The mass range used for analysis was *m/z* 45 - 200. The deconvolution of the test compound was performed using PARAFAC initiated by TLD [Sinha et al. 2004a; Sinha et al. 2004b]. The TLD algorithm was from the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA) and was selected for the advantageous sequencing of the three dimensions of the matrix prior to analysis. The PARAFAC algorithm was from the N-way Toolbox 2.01 [Andersson and Bro 2000]. Unimodal and non-negative constraints were applied to both chromatographic dimensions and non-negative constraints were applied to the mass spectral dimension. There were five components specified in the model, one of which accounted for baseline offset. Each PARAFAC analysis took ~ 30 seconds on a 1.6 GHz processor with 1024 MB of RAM.

6.3.3 AMDIS Analysis

Regions surrounding the test compound of the three replicate data sets for the environmental sample (1500 - 1560 s) were exported as ANDI or NetCDF (.cdf) files and loaded into AMDIS (Automated Mass Spectral Deconvolution and Identification System v 2.1; DTRA/NIST 2002) for data processing. The region between 25.305 and 25.373 minutes was processed for peak deconvolution. Since AMDIS analyzes the data in a GC-MS format, the 3

most abundant modulations of the test compound onto column 2 were analyzed. The mass range used for analysis was *m/z* 45 - 300. Five major parameters were set for AMDIS deconvolution. The component width equal to the number of scans across a well-resolved peak at half-height was set to the default of 12. Changing this value seemed to have little effect on the deconvolution results. The degree to which adjacent peaks are subtracted during deconvolution was set to two adjacent peaks. The resolution was set to high indicating a high level of overlap. The sensitivity was set to high to search for increasingly noisy and broad peaks. The shape requirements as the way of forcing all of the deconvoluted peaks (at each mass channel) to have the same shape (per component) were set to the default medium. Model peaks were produced for each of three column 2 modulations. The model that resulted in the best mass spectral similarity between the deconvoluted spectrum and the library spectrum for the test compound was determined to be the best model. Each AMDIS analysis took ~ 5 seconds on a 1.6 GHz processor with 1024 MB of RAM.

6.3.4 LECO Analysis

The LECO automated peak finding and deconvolution ChromaTOF software was used to analyze the region of environmental sample #2 that contained the test compound (1515 - 1540 s). A data-processing file was created to analyze this small region because analysis of the entire data set is overly time consuming (~ 3 hours). The data processing file consisted of the following elements: baseline, peak find, library search, and calculate area/height. Baseline tracking was 'on' for the entire data set with an offset of 0.5. The peak width was set to 0.1 s. The maximum number of unknown peaks to find was set to 100. Segmented processing was selected with peak find turned 'on' from 1515 - 1540 s. GC x GC parameters were applied with a

similarity match factor of at least 500 to combine multiple modulations per compound onto column 2. The peak width on column 1 was set as 5 s. The library identity search mode was normal (forward) and was set to return 10 hits. The maximum molecular weight was set to 300. Since all *m/z* channels (35 - 300) were used in the deconvolution analysis, the deconvoluted spectra were exported to NIST MS Search 2.0 where only *m/z* 45 - 300 were used for similarity matches. Each automated LECO analysis took ~ 30 seconds on a 2.2 GHz processor with 1024 MB of RAM. Manual re-processing of automated analysis added up to several minutes of processing time per analysis.

6.4 Results and Discussion

6.4.1 Environmental Sample

An environmental sample (#2), a mixture of fuels, natural products and pesticides, was analyzed with the GC x GC-TOFMS instrument generating a highly complex data set. The TIC of this data set is shown in Figure 6.1A. This image reflects only the most abundant components in the sample, as those are the most apparent in the TIC. Viewing other mass channels demonstrates the extensive complexity of the sample. With increasing complexity, there is a greater likelihood of peak overlap. One region in this chromatogram that exhibits overlap and contains the proprietary test compound spiked at 262 ng/ml is shown in Figure 6.1B. This complex region provides the data for the deconvolution comparison of PARAFAC, AMDIS, and the LECO ChromaTOF software.

6.4.2 PARAFAC Deconvolution Results

The region around the test compound that was analyzed with PARAFAC is shown as the sum of *m/z* 95 and 140 in Figure 6.1C. There are four labeled components visible. The PARAFAC deconvolution was initiated

TLD to provide the initial estimates of peak profiles. The analysis was performed in triplicate, but only one representative trial is shown for brevity. The PARAFAC deconvoluted column 1 and column 2 chromatographic profiles are shown in Figure 6.2A and B respectively (baseline component 5 omitted for clarity). Concentration information is contained within the column 2 profiles while the column 1 profiles are normalized. From Figure 6.2B, it can be seen that component 4 is about 8 times less abundant than component 1, the most prevalent component. Components 3 and 4 are also significantly overlapped on both columns, yet PARARFAC is able to deconvolute the signals.

The deconvoluted mass spectral profiles for components 1-3 were similarity matched with spectra in the NIST database to find their probable identities (not shown for brevity). Compounds 1 and 2 both matched to the same library spectrum of 2,6-dimethyl-naphthalene, but visually exhibited slight differences from each other indicating that they are most likely dimethyl-naphthalene isomers. The best similarity match for compound 3 was (E)-6-tridecen-4yne. The similarity match factors for these 3 components for all three trials were greater than 850. The PARAFAC deconvoluted mass spectrum for component 4 resulted in the highest similarity match with a compiled library spectrum of the test compound (not shown for brevity). The similarity match factor for the deconvoluted spectrum for this case was 798 with a reverse match factor of 801. The similarity and reverse match factors for all three replicates are listed in Table 6.1. Since the similarity match factors range from 735 - 798 (Table 6.1), some variability in the reproducibility of the deconvolution exists. While there is variability, it is manifested as variability in ion ratios, not the addition of contaminants into the spectrum. This type of variability would result in a much higher reverse match factors

compared to the similarity match factors. Since the reverse match factors are all within 3 points of the similarity match factors for the PARAFAC results (Table 6.1), it is evident that PARAFAC is not inserting contaminant *m/z* peaks into the deconvoluted spectra.

In order to compare these results to those of AMDIS and LECO, it is important to understand the data formatting for GC x GC data. The PARAFAC deconvoluted column 1 profiles (Figure 6.2A) are jagged because each data point represents a discreet modulation onto column 2 based on the modulation frequency and the width of the peak on column 1. For component 4, there are 5 distinct modulations as indicated by the abrupt changes in slope throughout the course of the profile. The three column 2 modulations with the most signal for component 4 are circled in Figure 6.2A and labeled **a**, **b**, and **c**, where **b** is the largest modulation. These three modulations form the basis of comparison for the AMDIS and LECO deconvolution.

6.4.3 AMDIS Deconvolution Results

Figure 6.3 shows the region of environmental sample #2 used in AMDIS analysis that contains the three most abundant modulations from column 1 onto column 2 of the test compound. The TIC (gray), *m/z* 156 (red - characteristic of dimethyl-naphthalenes), and *m/z* 113 (blue - characteristic of the test compound) are shown. This is the raw signal prior to formatting into a 3D array. Modulation locations are indicated by dashed lines and are labeled a - c. Between each dashed line is a complete column 2 chromatogram. When formatting the data into a 3D array, these individual column 2 chromatograms would be segmented off and then stacked adjacent to one another. The apexes of the three most abundant modulations of the test compound are marked with small black circles. These circles correspond to those in Figure 6.2A. Since AMDIS is intended for use with GC-MS data, the

three modulations for the test compound were each analyzed separately for the three replicate samples. The apexes of three modulations (**a** - **c**) of the test compound occurred at 25.311 min, 25.336 min, and 25.361 min. It should be noted that the units on the time axis drawn by the AMDIS software are in minutes, which for the data acquisition rate of the TOF, do not have enough significant figures to accurately label the axis. As can be seen in Figure 6.4A, as many as 6 tick marks correspond to the label 25.311.

For each modulation of the test compound onto column 2, the deconvolution program selects unique ions from the apexing compound and unique ions for up to two interfering compounds. Deconvolution models are generated from a peak model for the apexing compound based on its own unique ion(s) (+*m/z*) and from subtracting information associated with the unique ions from interfering compounds (-*m/z*). Multiple models are generated for each modulation onto column 2. The models that exhibited the best mass spectral similarity match for the test compound were considered the best models and are shown for modulations **a** - **c** in Figures 6.4A - C. The analyses were performed in triplicate, but only one representative analysis is described here for brevity. For modulation **a** at 25.311 min (Figure 6.4A), model 5 (+113, -156) of 13 possible resulted in the best deconvoluted spectrum with a similarity match factor of 691 (Reverse = 695) (Table 6.1). For modulation **b** at 25.336 min (Figure 6.4B), model 1 (+57, -156) of 1 possible resulted in the best deconvoluted spectrum with a similarity match factor of 793 (Reverse = 796) (Table 6.1). For modulation **c** at 25.361 min (Figure 6.4C), model 2 (+113, -156) of 6 possible resulted in the best deconvoluted spectrum with a similarity match factor of 770 (Reverse = 774) (Table 6.1). The similarity and reverse match factors for the test compound of the two additional replicate data sets for each modulated peak are listed in Table 6.1 according to

replicate number and modulation **a** - **c**. Variability is evident for the match factors between trials and between modulations (Table 6.1). Modulation **b** generally has the best match factors due to its increased S/N as the largest modulation. As for the PARAFAC results, the similarity and reverse match factors agree well with each other, indicating that the deconvolution is not adding erroneous *m/z* peaks into the deconvoluted mass spectra (Table 6.1). As can be seen from the TIC (gray) in Figure 6.4A and Figure 6.4C, there is at least one other component overlapped with the test compound at 25.311-25.312 min and 25.361-25.362 respectively. The deconvolution models, however, do not reflect the additional interference. Even though two interfering peaks were allowed, the program only subtracted one interfering peak model corresponding to *m/z* 156 for modulations **a** - **c**.

6.4.4 LECO Deconvolution Results

The LECO deconvolution algorithm is proprietary, but it processes the data in a way similar to AMDIS by treating the data set as one long GC-MS chromatogram. Theoretically, all the peaks in the data analysis region are found, deconvoluted, and library matched. Peaks from adjacent column 2 modulations with spectral similarities greater than 500 to one another are then combined as one GC x GC peak. The peak labels generated after GC x GC grouping conform to a standard format. For example, peak 1 has three modulations onto column 2 resulting in three column 2 chromatograms. For each modulation there is a peak labeled 1-1, 1-B, and 1-3 respectively, where 1-B denotes the base peak. The results of the LECO automated peak finding and deconvolution analysis are shown in Figure 6.5A. Dashed lines indicate modulation or column 2 modulation locations. For Figure 6.5A, modulations onto column 2 and the complete column 2 chromatograms between the dashed lines are labeled **a** - **c**.

The test compound was not located in the automated analysis performed. As can be seen from the chromatogram for modulation **a** (Figure 6.5A), the peak labeled 3-3 is actually in a location occupied by three distinct peaks. One possible explanation for this is the data acquisition frequency. The rate that spectra were collected for these data sets was optimized based on the guideline of 10-12 points (i.e., full scan spectra) across a column 2 peak for adequate peak shapes, minimizing the size of the data files (fewer data points), and maximizing S/N (more signal averaging). The optimal number of data points across a peak for the LECO ChromaTOF software is 20 - 30, which is more than were acquired for this sample set.

By manually deleting peaks 3 and 4 (4-B and 4-2 are column 2 modulations of the same compound where 4-B is the base peak) in Figure 6.5A and manually adding and appropriately grouping new peaks 3, 4, and 5 in Figure 6.5B, the ChromaTOF software can be coaxed into generating an acceptable deconvolution for the test compound (peak 4). Information on the expected retention time and spectral features of the test compound were necessary to generate the desired results. The same three column 2 modulations and subsequent separations for the test compound that are circled and labeled in Figures 6.2A and 6.3 are shown in Figure 6.5B. The three column 2 peaks for the test compound are labeled 4-1, 4-B, and 4-3. During manual reprocessing for modulation **a** (Figure 6.5B), it was important to add the peaks in a certain order (3, 4, then 5), otherwise the three compounds including the test compound and the two interfering analytes resulted in contaminated deconvoluted spectra. The three deconvoluted peak profiles for the test compound (peak 4) are shown in Figures 6.6A - C. The *m/z* selected for each deconvoluted compound in the figures is due to the

unique ion used for deconvolution. For the test compound, *m/z* 95 was the selected ion and provides the peak profiles (blue lines) in Figures 6.6A - C.

The deconvoluted mass spectra for the test compound (peak 4) and the two major interfering peaks (3 and 5) were identified based on the best similarity matched library spectra (not shown for brevity). The deconvoluted spectra for each of the test compound peaks (4-1, 4-B, and 4-3) show some differences (not shown for brevity). The spectrum for the base peak, 4-B, is given as the true deconvoluted spectrum due to the increased S/N of this modulation compared to the smaller modulations for the compound. The resulting similarity match factor for this spectrum with the compiled library spectrum for the test compound was 815 (Table 6.1). For the other two trials, the similarity match factors were 828 and 833; however, the reverse match factors for all three trials were on average 12 points higher than the similarity match factors (Table 6.1). Although the differences between the similarity and reverse match factors for the LECO ChromaTOF deconvolution were higher than those for PARAFAC and AMDIS, they still indicate an acceptably low level of contaminant *m/z* peaks that do not appear in the library spectrum. Peak 5 had a high similarity match (> 900) to 2,6-dimethyl-naphthalene for all three replicate trials. The deconvolution of peak 3 showed a much lower similarity match factor (740) to the best match compound, 1,2-dimethyl-1,3-cyclopentadiene, but a rather high reverse match (919) indicating significant contamination of the spectrum. The other two trials for peak 3 resulted in best matches of 1-methyl-1,4 cyclohexadiene (similarity: 746; reverse: 922) and (E,E)-1,3,5-heptatriene (similarity: 729; reverse: 876) confirming the suspect deconvolution results for this component.

6.4.5 Deconvolution Methods Comparison

Both PARAFAC and AMDIS analyses currently require exporting data from the LECO ChromaTOF software, which adds an additional step for analysis. With the high complexity of this data, exporting an entire data set can take a substantial amount of time (often ~ 30 minutes). Since deconvolution analysis regions are very small compared to the size of an entire data set, only small portions of the data need to be exported for analyses on AMDIS or PARAFAC. This reduces the time expenditures for exporting and the use of disk space to hold large exported complete data files (~ 1 GB for a 45 minute GC x GC-TOFMS separation in .csv format). The built-in nature of the LECO ChromaTOF software is a strong benefit.

With data files of this magnitude, a deconvolution program that provides an automated analysis system is highly desirable. Both LECO and AMDIS are automated, while PARAFAC is a manual technique. In addition, the PARAFAC analysis also requires familiarity with the Matlab programming environment and language. This can be an initial deterrent to using this technique; however, the number of Matlab commands actually required to manipulate the data sets, perform the deconvolution, and evaluate the results is relatively small. With this particular data set the LECO ChromaTOF results required some additional manual data processing to obtain the desired deconvolution results. This added time and effort to the deconvolution process. In addition, without *a priori* knowledge of the test compound's expected retention time or spectral features, the test compound would have been overlooked using the automated software. With a greater number of full scan spectra across each column 2 peak, this process might have been reduced, but the signal for small analytes would have been greatly diminished. There appears to be a trade-off in the LECO ChromaTOF

software for signal height of small analytes and automated deconvolution results. The AMDIS analysis is automated, but there is some additional manual processing required locating the multiple column 2 separations for each component since the program was designed for GC-MS data. Ultimately, the amount of time needed for complete analysis of the submatrix for this study was roughly comparable for each deconvolution technique.

Both AMDIS and the LECO ChromaTOF software generate column 2 peak profiles and mass spectra for deconvoluted components, while PARAFAC provides column 1 peak profiles as well. Depending on the unique ions found by AMDIS and the LECO ChromaTOF software for an analyte and its interfering neighbors, the column 2 profiles may be generated from peak models based on different ions. This could lead to some variations in column 2 peak shapes across the multiple column 2 separations of an analyte. Since PARAFAC treats the data as a whole, there is significantly less ambiguity in determining the entire signal corresponding to one particular analyte. PARAFAC does not require a regrouping step after analysis, unlike AMDIS and LECO, potentially resulting in more consistent results. PARAFAC is also the only deconvolution technique tested here that can provide enough information to reconstruct the region of analysis showing signal only from a desired analyte. This is done for one analyte by taking the outer product of the pure component profiles. After reconstruction, the concentration information included in the column 2 profiles provides the information necessary to quantify the analyte as though it was completely chromatographically separated initially.

While all three deconvolution techniques provided spectra that enabled the correct identification of the test compound, the potential for quantitation among these three techniques varies. With PARAFAC, manual quantitation is

straightforward and accurate provided there is acceptable chromatographic and mass spectral selectivity in the data [Sinha et al. 2004b]. AMDIS was developed to deconvolute mass spectra for analyte identification and structure analysis without an emphasis on concentration quantitation. While the AMDIS deconvolution results do provide peak widths, purity, abundances, amounts, S/N (total), and total signal for each column 2 separation of an analyte, the method one would use to recombine the information from the deconvolution models to form and quantify one analyte signal is not readily apparent. The LECO ChromaTOF software provides an integrated, quantitative software tool that generates the S/N, area, and height of an analyte signal based on either the unique ion, or a user specified ion. Provided that the software is able to successfully find and deconvolute all of the column 2 separation peaks for one analyte and group them correctly, quantitation is possible and largely automated.

6.5 Summary and Conclusions

Although not a comprehensive comparison of deconvolution techniques, some useful insights were presented in this study regarding the applicability of currently available peak deconvolution methods to GC x GC-TOFMS data. There are benefits and drawbacks to each deconvolution method discussed.

PARAFAC is currently a manual technique that requires some selectivity in each dimension to successfully deconvolute a region of interest. This technique also currently requires exporting data and analyzing it with Matlab. PARAFAC is able to analyze the data as a whole and provide pure component peak profiles in each dimension that are suitable for straightforward analyte identification and quantitation. The number of

overlapping signals in the analysis region for PARAFAC is limited by the size of the submatrix, but typically has an upper limit of five or six.

AMDIS is an automated technique that relies on the presence of a selective ion in the analyte of interest and in each interfering component to generate peak shape models for deconvolution. This technique requires exporting data and analyzing it in the AMDIS software program. AMDIS is not able to view GC x GC-TOFMS data as a whole because it was designed for use with single dimension GC-MS. This necessitates an additional manual-processing step to recombine all the information from one compound. AMDIS is also primarily useful for identification with only limited quantitative abilities. In addition, only two interfering components can be subtracted from an analyte of interest during deconvolution.

The LECO ChromaTOF software provides an automated deconvolution technique that also relies on the presence of a selective ion in the analyte of interest and in each interfering component to generate peak shape models for deconvolution. This technique is integrated into the software used to run the instrument, thus requiring no exportation of data. While the LECO software still processes the GC x GC-TOFMS data as a GC-MS chromatogram, it is designed to automatically regroup all of the column 2 separations corresponding to one analyte after deconvolution. This only works if the software is actually able to locate and deconvolute all of the present peaks. Quantitative tools are built into the software, usually providing quantitation based on the unique ion found during deconvolution. The number of interfering compounds capable of being deconvoluted from an analyte of interest is not limited to two, as for AMDIS. In the case presented, the LECO ChromaTOF software was not able to detect, deconvolute, and quantitate the

test compound in a completely automated fashion. The processing required manual intervention to obtain the desired results.

While no one deconvolution method studied here provides the perfect solution to signal deconvolution in GC x GC-TOFMS data, the potential is great for an improved method that integrates features from all of the methods discussed. Ideally a data analysis and peak deconvolution system will be automated, treat GC x GC-TOFMS data as a whole, and provide accurate and quantitative information.

Table 6.1. Comparison of Mass Spectral Match Factors. Mass spectral similarity and reverse match factors for the deconvoluted test compound compared to a library spectrum for the test compound. Results are reported for comparison of PARAFAC, LECO ChromaTOF, and AMDIS deconvolution results. PARAFAC and LECO ChromaTOF results are listed for each of three replicate analyses. The AMDIS results are listed for each modulation (a - c) for each of three replicates.

Method and Replicate #	Similarity Match Factor	Reverse Match Factor
PARAFAC		
1	798	801
2	735	738
3	773	775
AMDIS		
1a	691	695
1b	793	796
1c	770	774
2a	724	732
2b	790	794
2c	751	757
3a	753	759
3b	805	807
3c	710	713
LECO ChromaTOF		
1	815	824
2	828	843
3	833	847

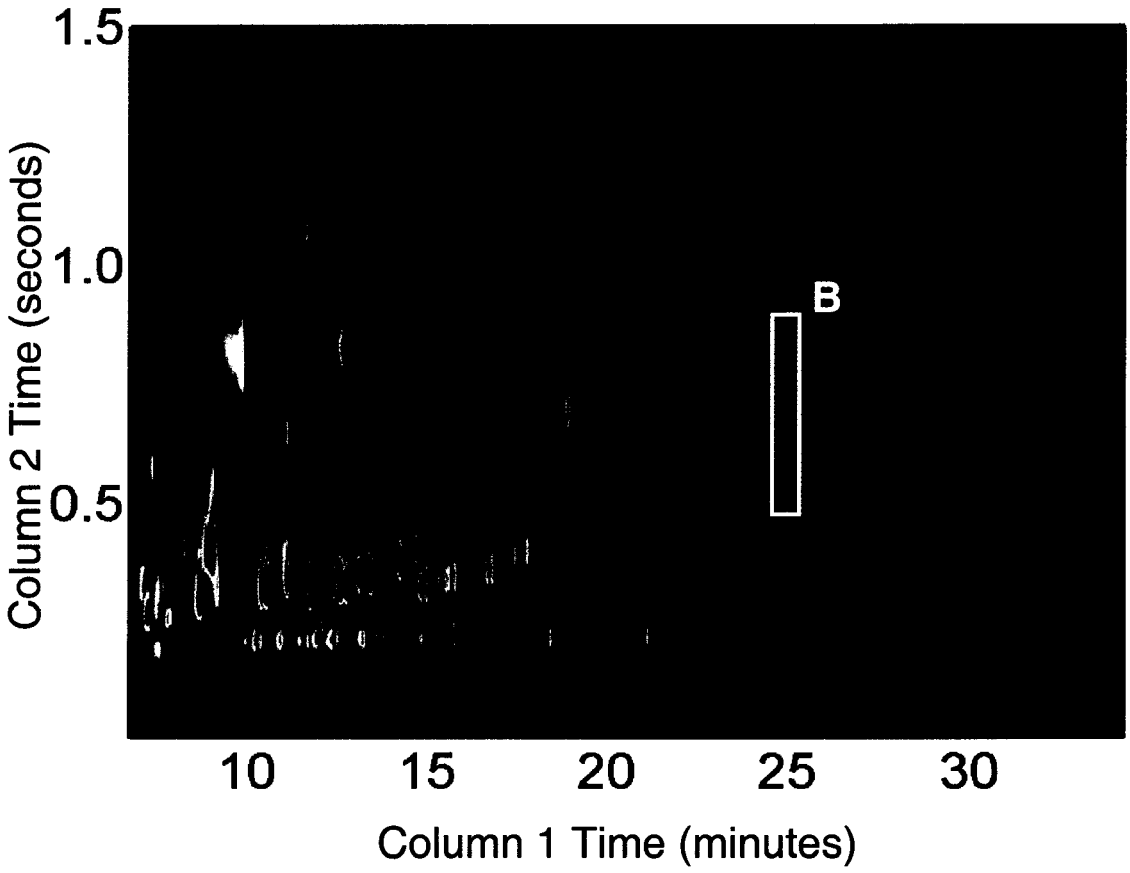


Figure 6.1A. Total ion current (TIC) chromatogram for environmental sample #2 containing fuels, natural products and pesticides. Area outlined in box is shown in Figure 6.1B. Note that the colormap is scaled according to the largest components in the region. Selecting a smaller region rescales the colormap so that peaks not readily visible in Figure 6.1A are apparent in Figure 6.1B.

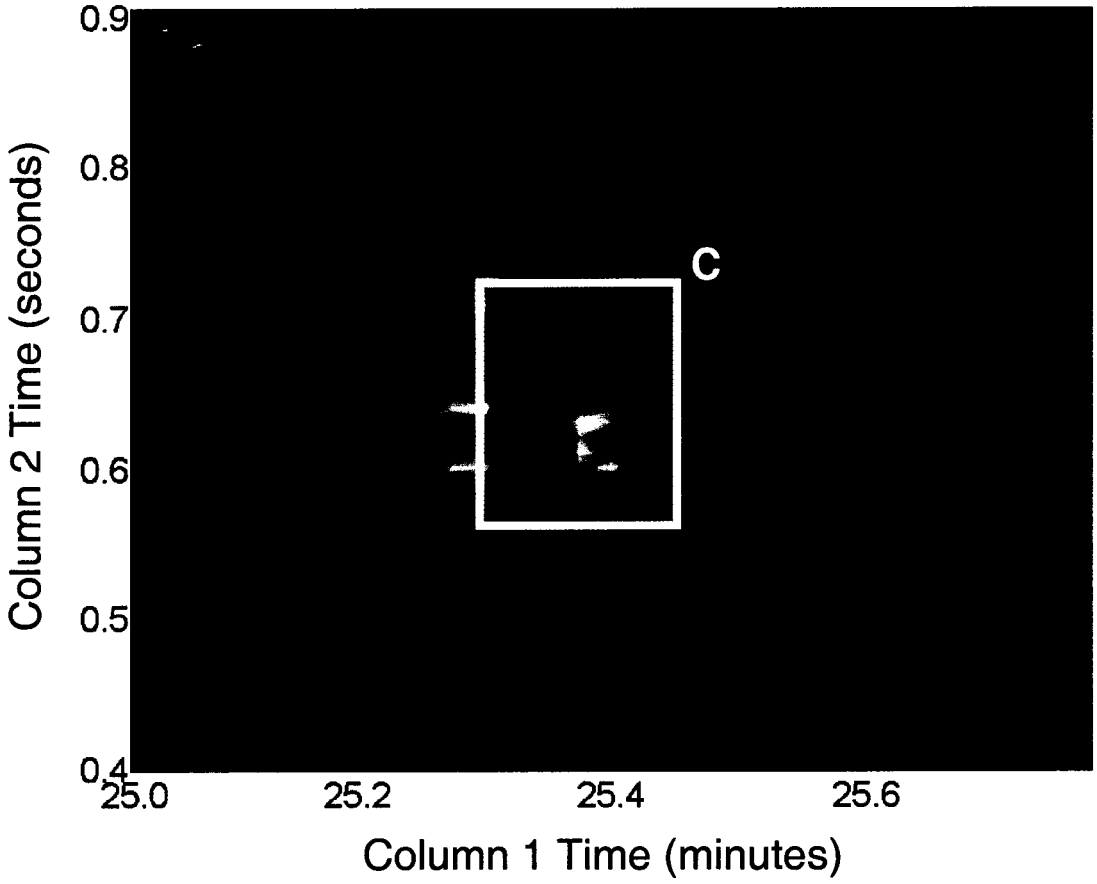


Figure 6.1B. Sub-region of TIC (*m/z* 40 – 200) chromatogram of environmental sample #2 showing the region of interest for deconvolution analysis. Region outlined in box was analyzed by PARAFAC and is shown in Figure 6.1C.

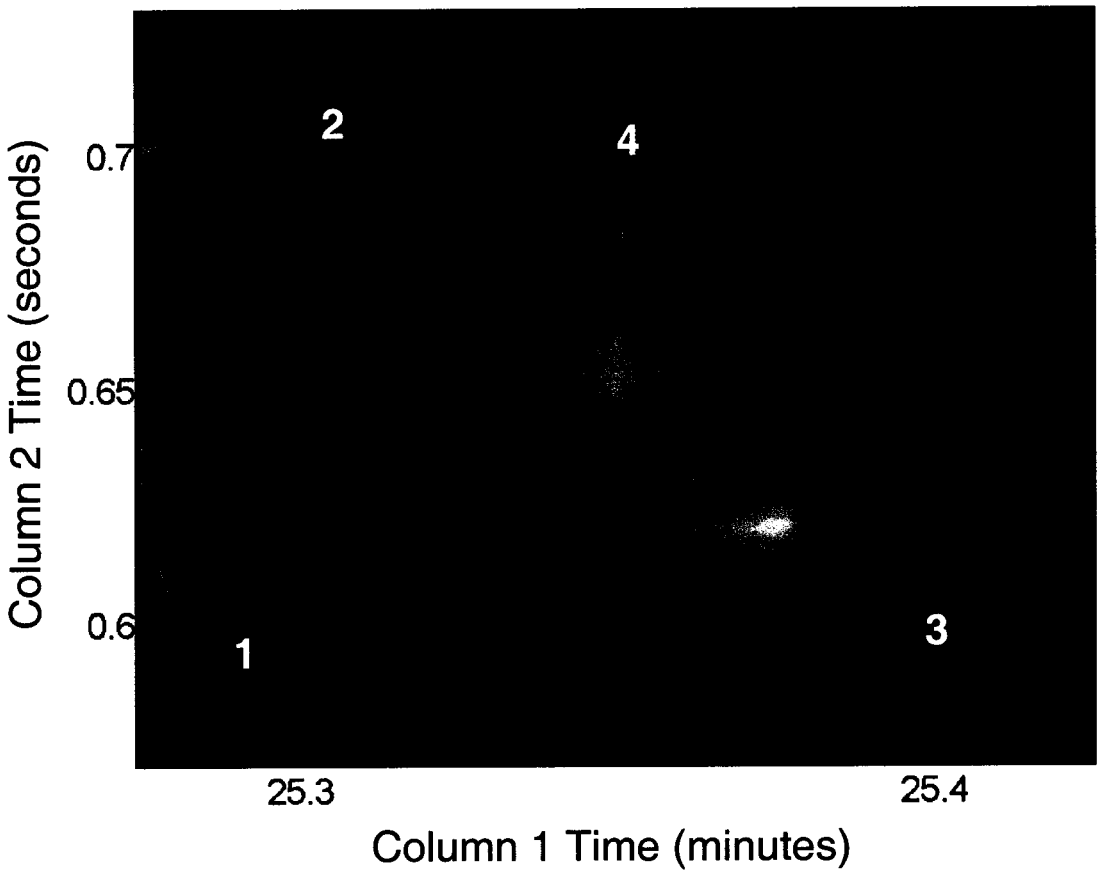


Figure 6.1C. Chromatogram of the sum of *m/z* 95 and 140 to demonstrate the four components in this region. This region was subjected to PARAFAC analysis.

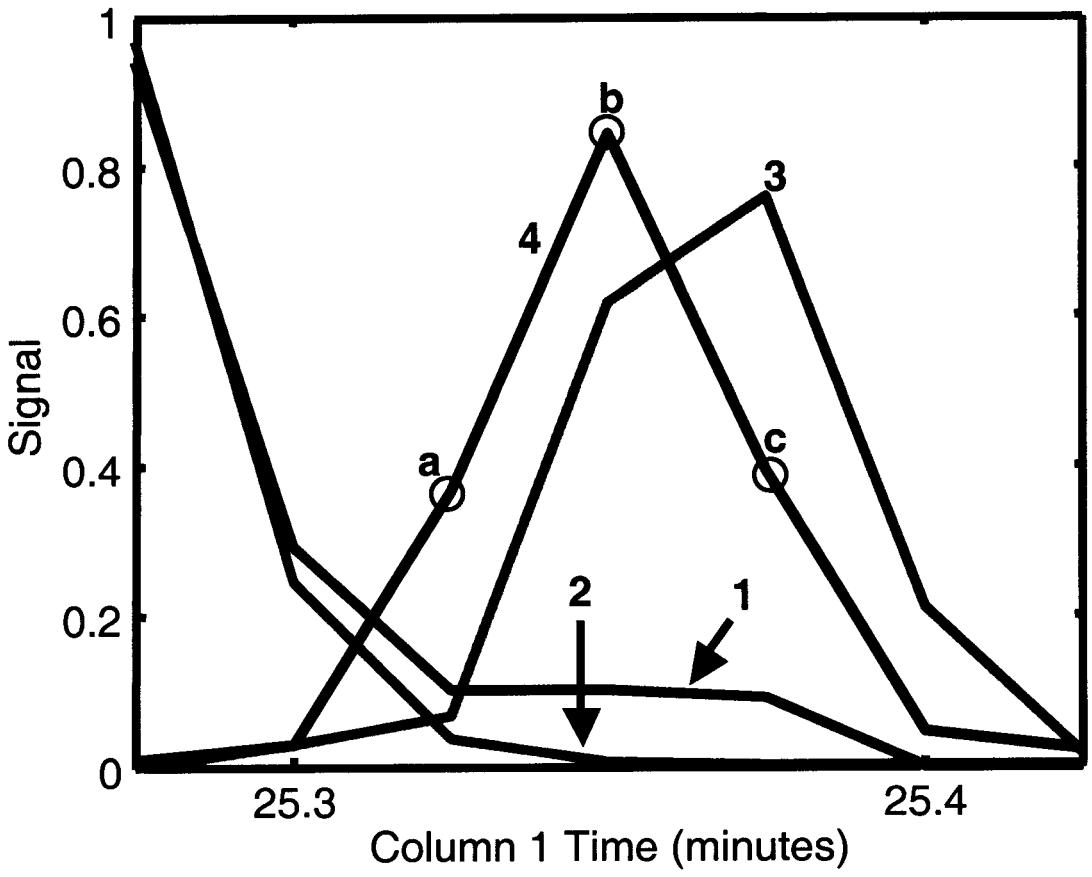


Figure 6.2A. PARAFAC deconvoluted column 1 profiles for the region shown in Figure 6.1C. The jagged profiles are due to the sampling frequency on column 1. At each modulation point there is a column 2 chromatogram. The three modulations onto column 2 that are used for comparison analyses with AMDIS and LECO ChromaTOF software are circled and labeled on component **4** (the test compound) as **a**, **b**, and **c**.

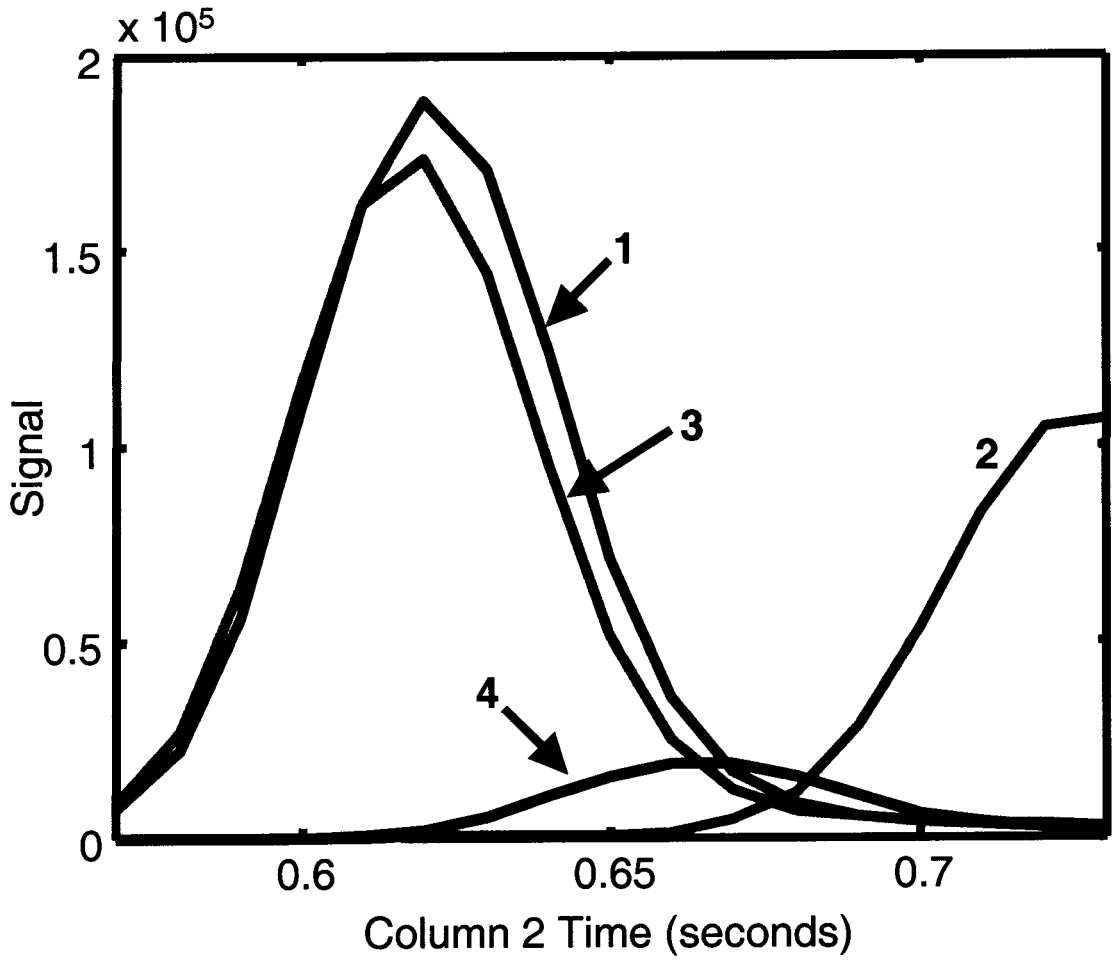


Figure 6.2B. PARAFAC deconvoluted column 2 profiles for the region shown in Figure 6.1C. Concentration information is included in the column 2 profiles.

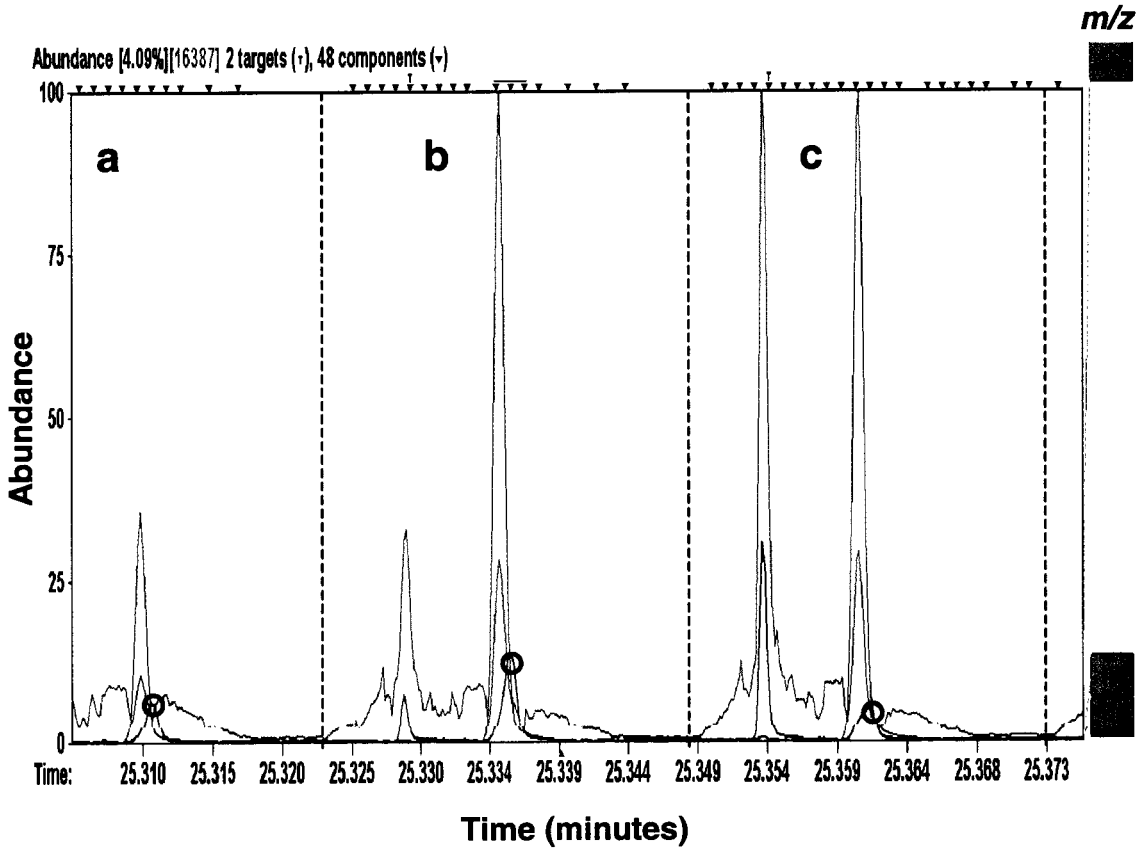


Figure 6.3. Region of environmental sample #2 used in AMDIS analysis that contains the three most abundant modulations from column 1 onto column 2 of the test compound. The TIC (gray), *m/z* 156 (red), and *m/z* 113 (blue) are shown. This is the raw signal prior to formatting into a 3D array. Modulation locations are indicated by dashed lines. Between each dashed line is a complete column 2 chromatogram labeled **a** – **c**. When formatting the data into a 3D array, these individual column 2 chromatograms would be segmented along the dotted lines and then stacked adjacent to one another. The apexes of the three most abundant modulations of the test compound are marked with small black circles. These circles correspond to those in Figure 6.2A.

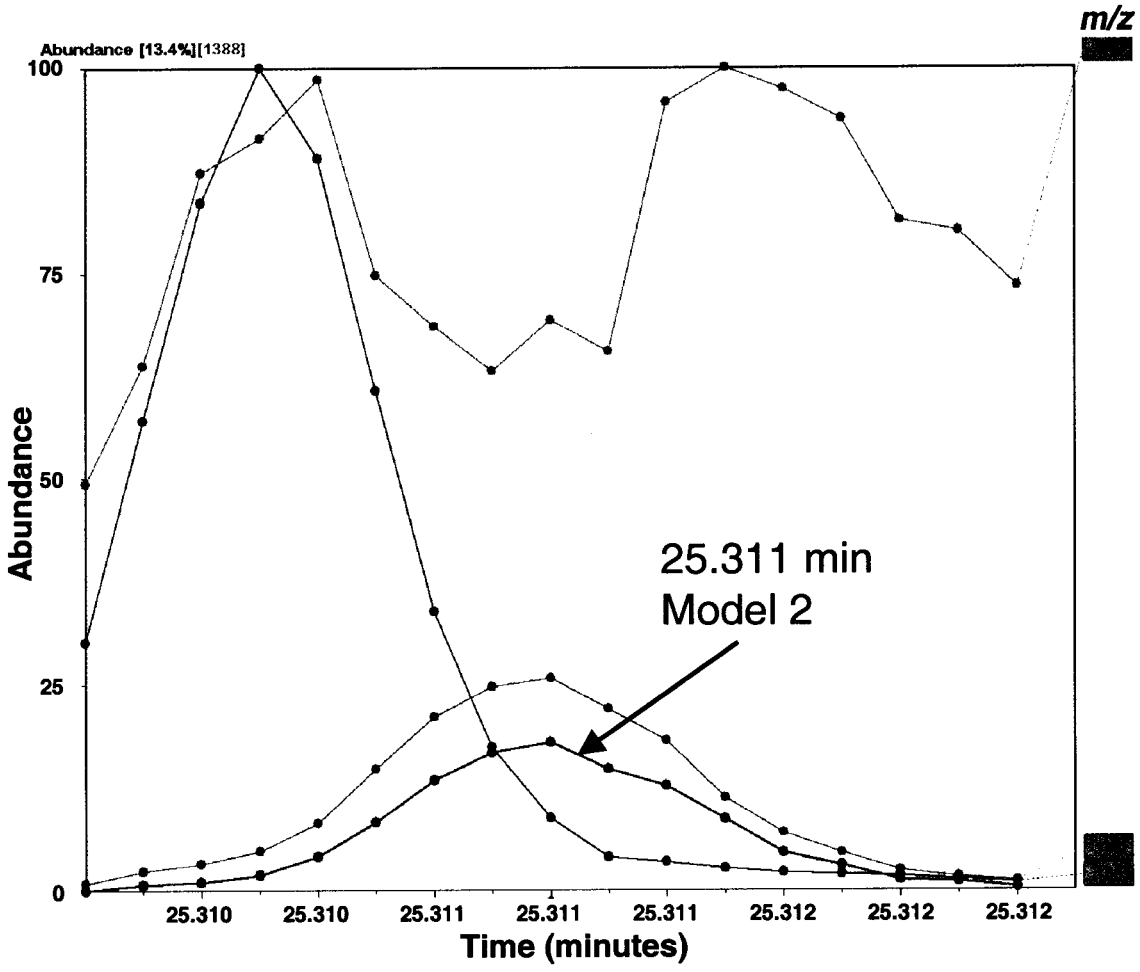


Figure 6.4A. AMDIS deconvolution profiles for modulation **a** of the test compound (R_t = 25.311 min). The TIC is shown in gray, *m/z* 113 in red, *m/z* 156 in blue and the deconvoluted profile in black. Model 2 is shown due to the best similarity match of the deconvoluted spectrum to that of the library spectrum for the test compound.

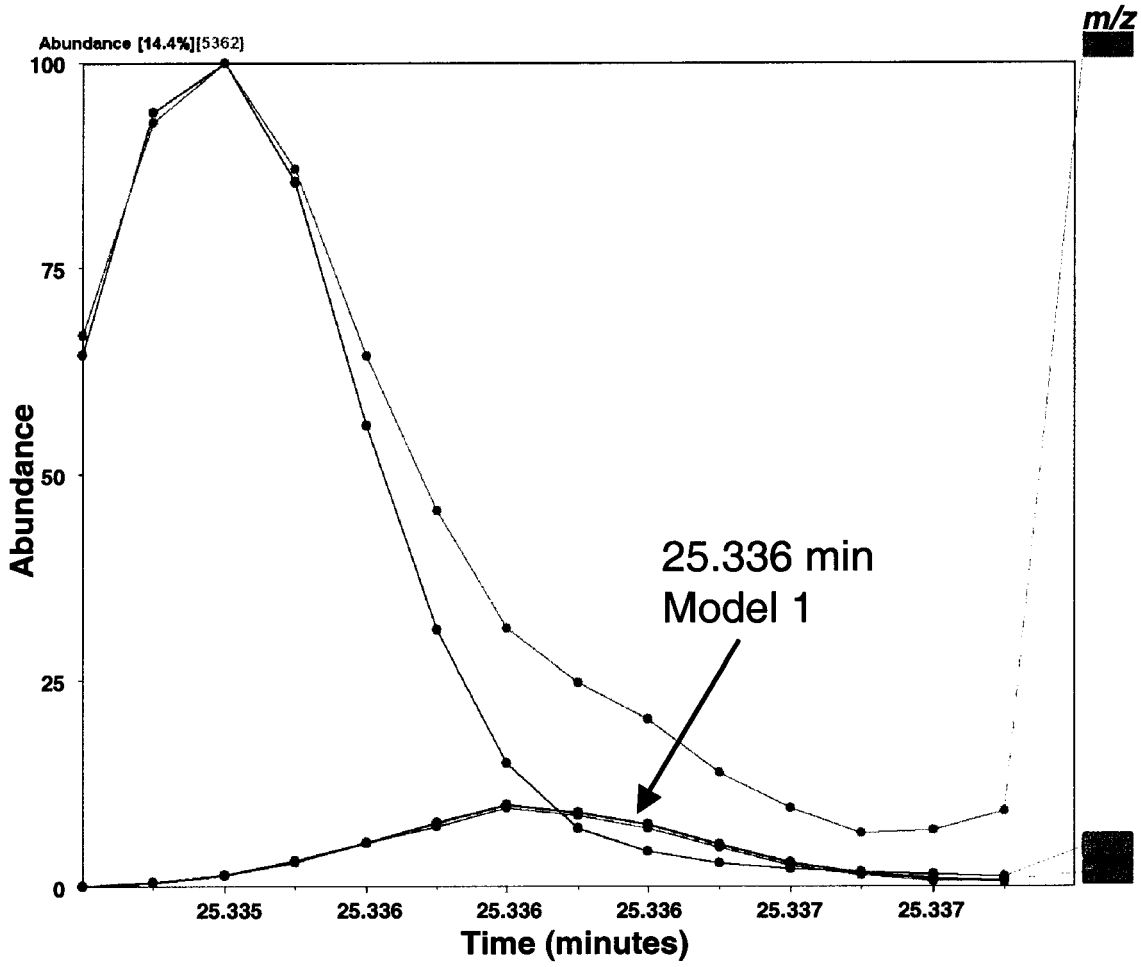


Figure 6.4B. AMDIS deconvolution profiles for modulation **b** of the test compound (R_t = 25.336 min). The TIC is shown in gray, *m/z* 57 in red, *m/z* 156 in blue and the deconvoluted profile in black. Model 1 is shown due to the best similarity match of the deconvoluted spectrum to that of the library spectrum for the test compound.

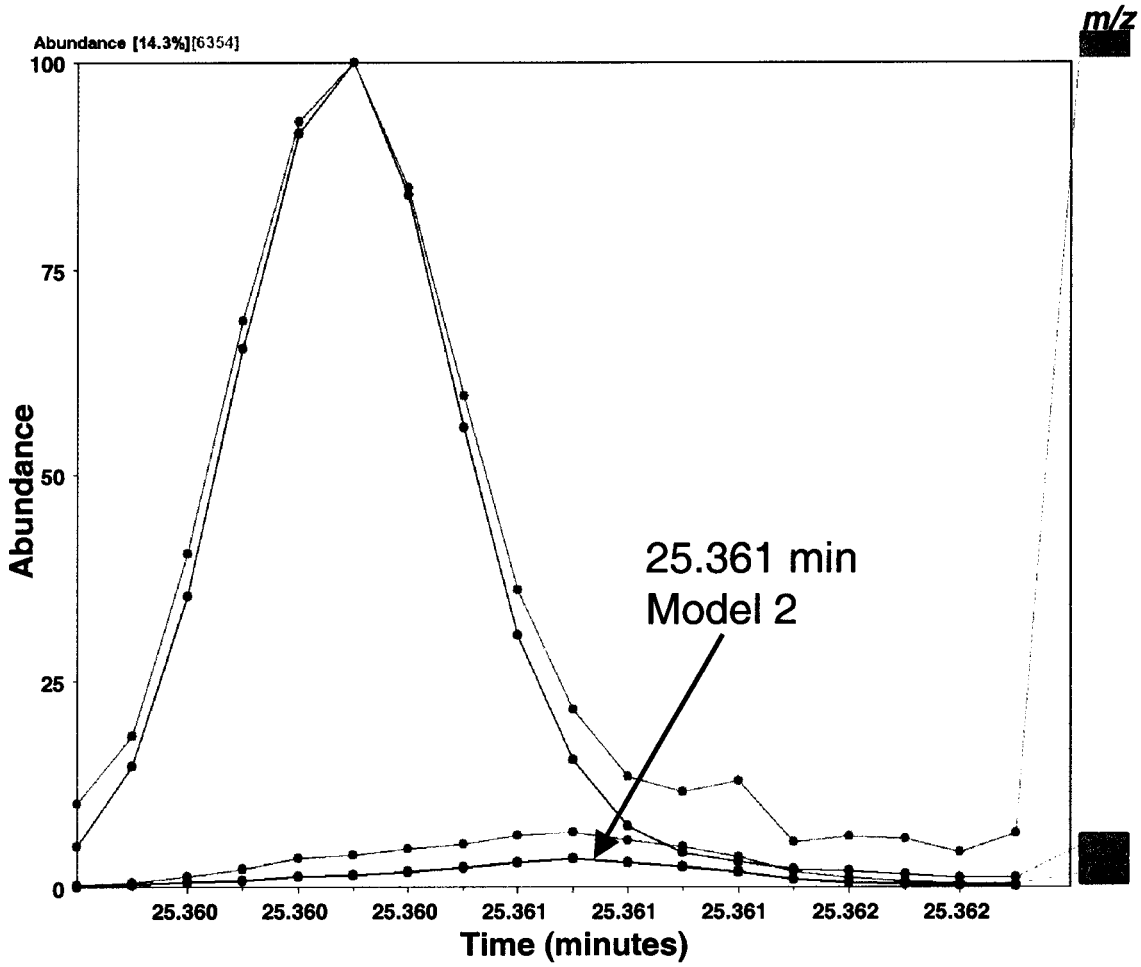


Figure 6.4C. AMDIS deconvolution profiles for modulation **c** of the test compound (R_t = 25.361 min). The TIC is shown in gray, *m/z* 113 in red, *m/z* 156 in blue and the deconvoluted profile in black. Model 2 is shown due to the best similarity match of the deconvoluted spectrum to that of the library spectrum for the test compound.

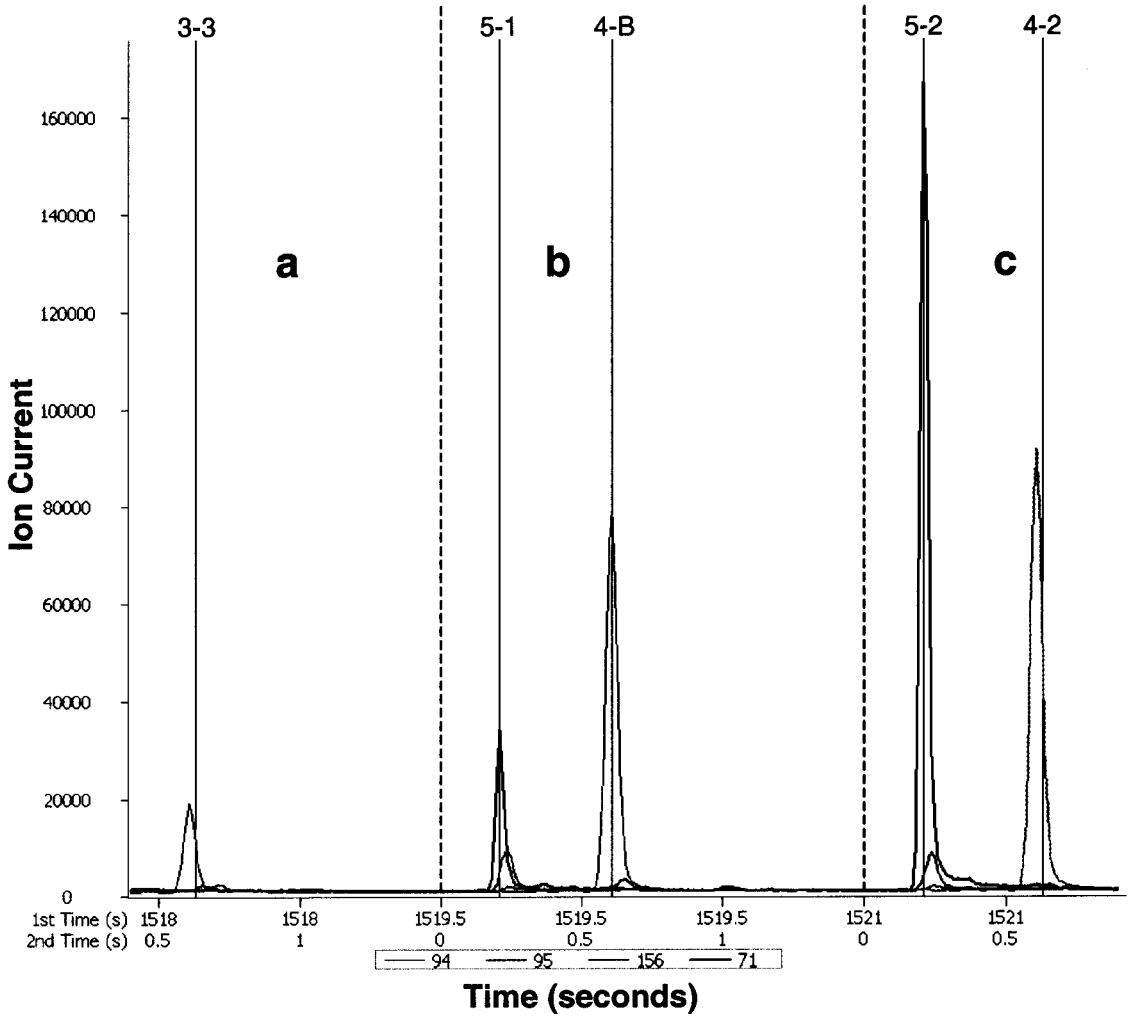


Figure 6.5A. LECO software results for automated peak finding and deconvolution. The test compound was not found and deconvoluted automatically. Modulation locations are indicated by dashed lines. Between each dashed line is a complete column 2 chromatogram labeled **a** – **c**.

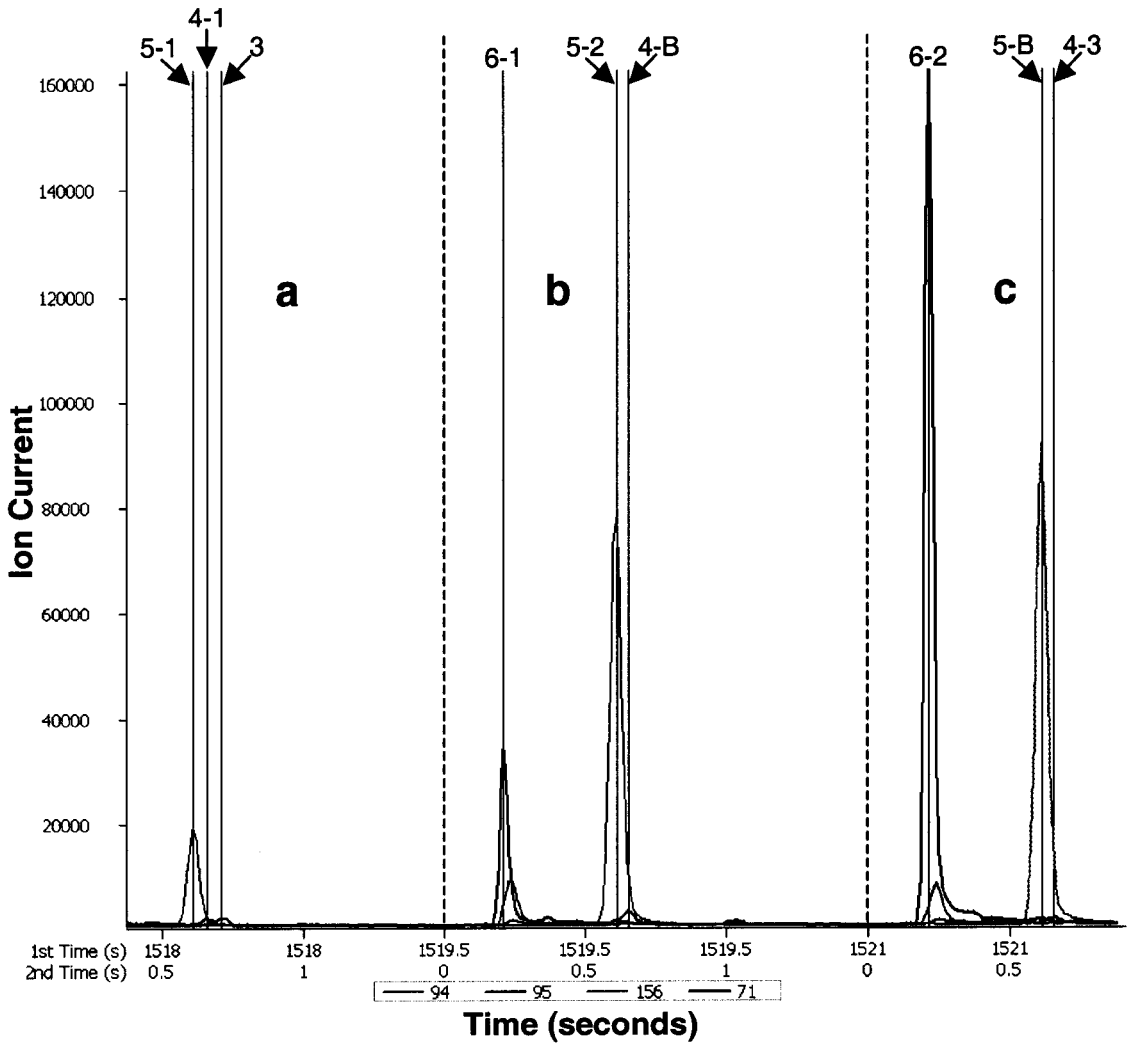


Figure 6.5B. LECO deconvolution result with manual re-processing. The test compound is peak 4 with the 3 modulations labeled **4-1**, **4-B**, and **4-3**. **4-B** denotes the base peak. Modulation locations are indicated by dashed lines. Between each dashed line is a complete column 2 chromatogram and labeled **a – c**.

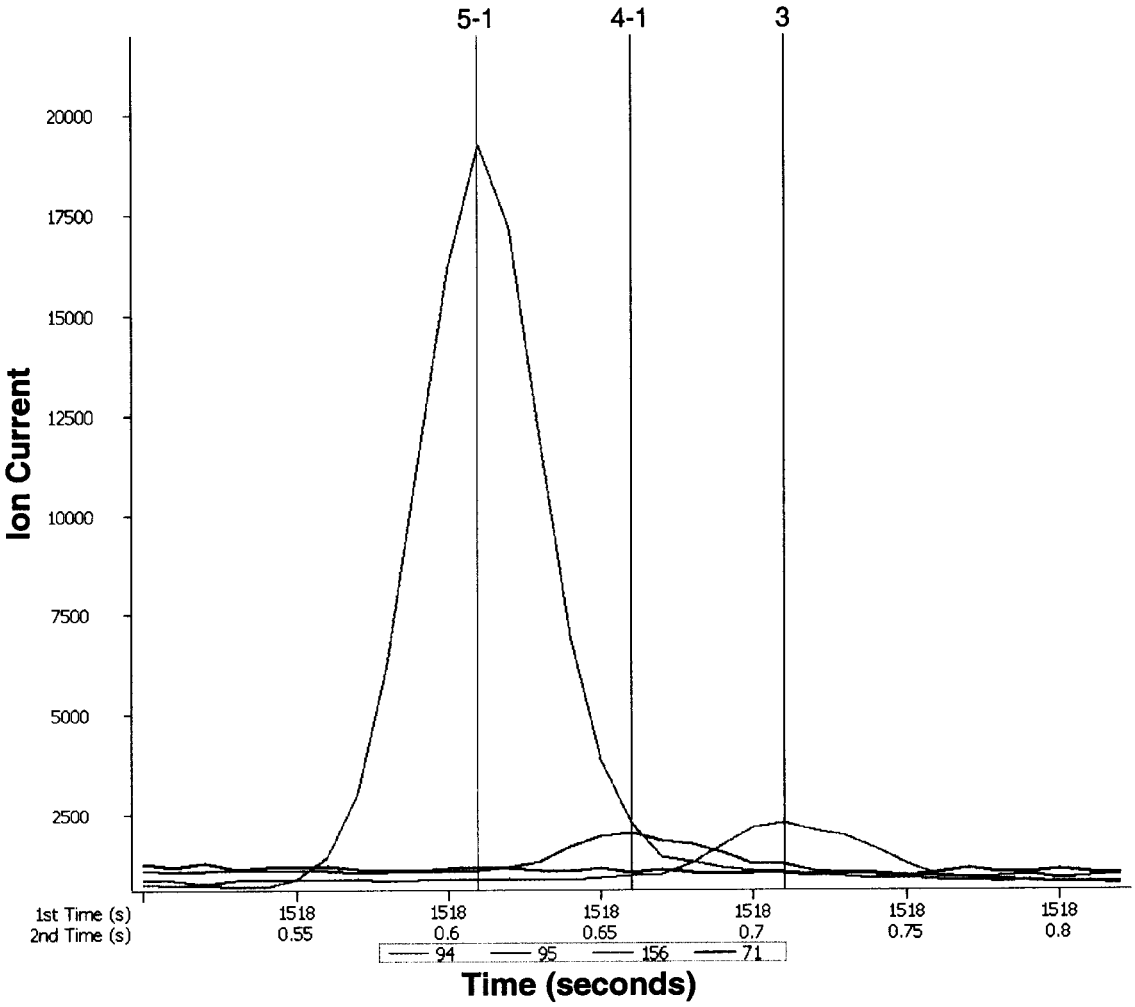


Figure 6.6A. LECO deconvolution with manual re-processing for modulation **a**. The test compound is peak **4-1** (blue peak trace). There are clearly two overlapping components (**5-1** and **3**).

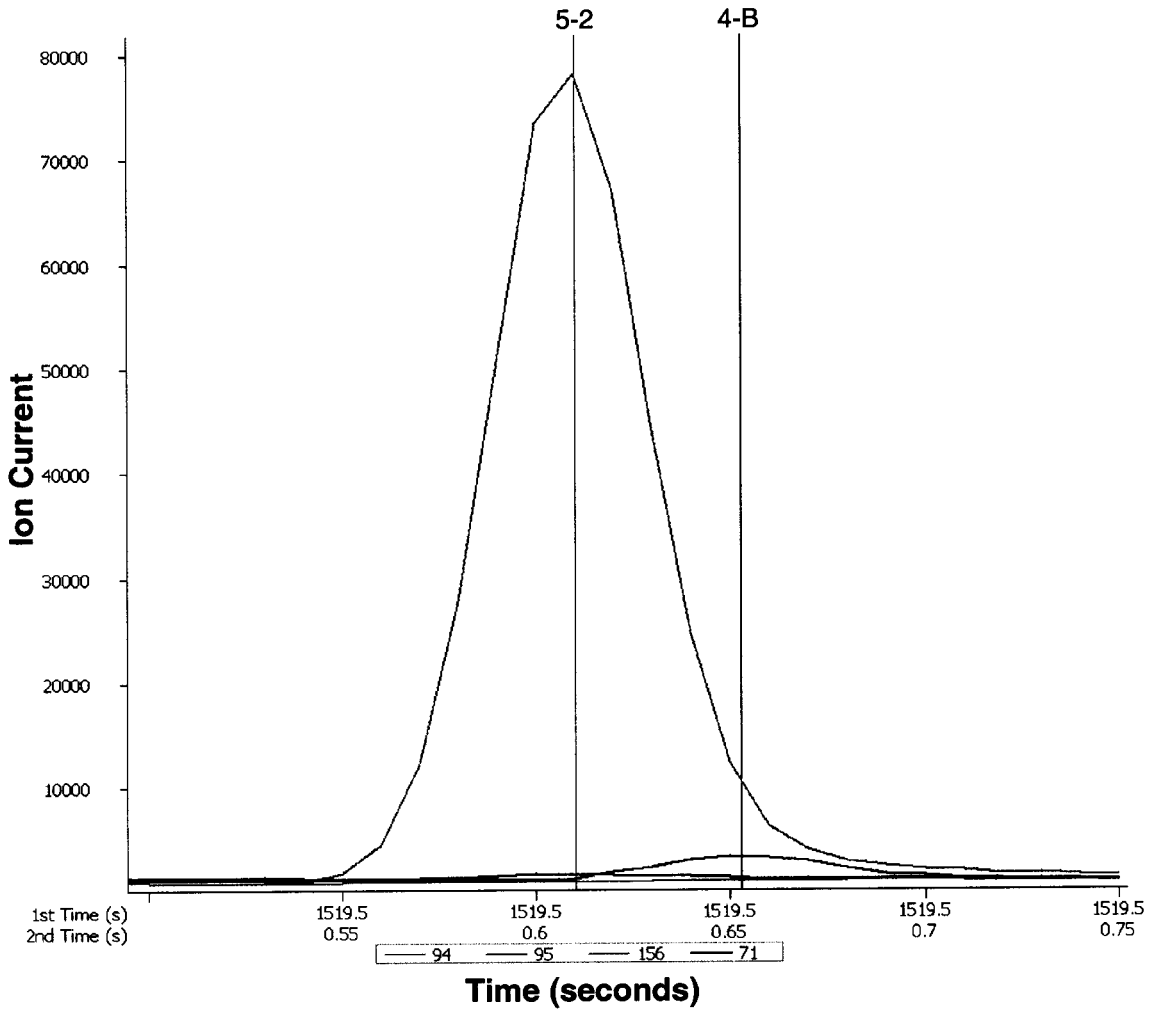


Figure 6.6B. LECO deconvolution with manual re-processing for modulation **b**. The test compound is peak **4-B** (blue peak trace).

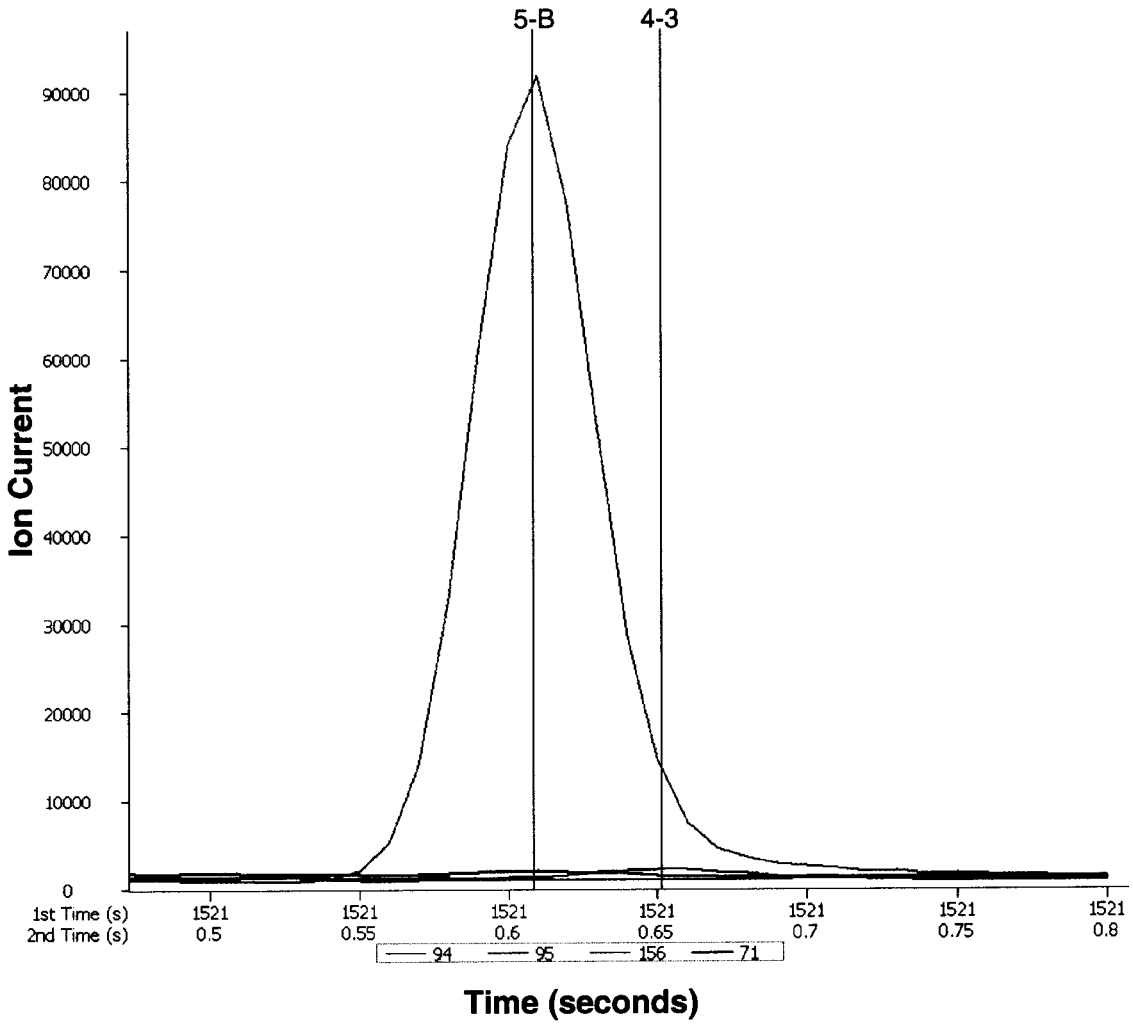


Figure 6.6C. LECO deconvolution with manual re-processing for modulation **c**. The test compound is peak **4-3** (blue peak trace).

CHAPTER 7. An Algorithm for Locating Analytes of Interest based on Mass Spectral Similarity in GC x GC - TOFMS Data: Analysis of Metabolites in Human Infant Urine*

7.1 Introduction

Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry (GC x GC-TOFMS) has been demonstrated as a highly selective instrument well designed for complex mixture analysis [Adahchour et al. 2003; Dallüge et al. 2002a; Dallüge et al. 2002b; Dallüge et al. 2002c; Dimandja 2003; Focant et al. 2003; Korytar et al. 2003; Lu et al. 2003; Shellie et al. 2001b; Sinha et al. 2003b; Sinha et al. 2004a; van Deursen et al. 2000; Welthagen et al. 2003]. Recently reported complex sample analyses include cigarette smoke [Dallüge et al. 2002b; Lu et al. 2003], petrochemicals [van Deursen et al. 2000], pesticide residues [Dallüge et al. 2002a; Zrostlikova et al. 2003], halogenated compounds [Focant et al. 2003; Korytar et al. 2003], airborne particulate matter [Welthagen et al. 2003], and flavors and fragrances [Adahchour et al. 2003]. The amount of data generated by this system is at best unwieldy and at worst can be nearly impossible to handle. Therefore the wealth of the data is immense; however, only after the appropriate data analysis tools are developed to extract the significant information can the value of the data be more fully realized.

For many applications involving complex mixture analysis, there is a need to locate and quantify specific analytes of interest [Harju et al. 2003a;

---

* Large portions of this Chapter have been previously published [Sinha et al. 2004c].

Harju et al. 2003b; Hua et al. 2003; Truong et al. 2003; van Mispelaar et al. 2003; Zrostlikova et al. 2003]. Often, matrix induced retention time shifts make analyte peak locations inconsistent from one analysis to the next [Dietz 1996; Fraga et al. 2000b; Fraga et al. 2001a; Poe and Rutan 1993]. In the currently available software, analyte peak location information is only available after processing an entire data set, which can be excessively time consuming. This typically involves generating a list of analyte peaks for a sample and matching as many as possible to library spectra. Although this is an automated process, it can sometimes take hours to complete. It is not efficient to process an entire data set of a complex matrix where there are perhaps thousands of compounds, when only a smaller subset of compounds are of interest.

These reasons constitute the impetus for the algorithm called "DotMap" reported herein. The DotMap algorithm employs a spectrum of interest and a complete GC x GC-TOFMS data set to search for the desired analyte in the complex data matrix. This spectrum of interest can be from a database such as NIST02 or from a user's collected library. After searching for the spectrum in the complete data set, a contour plot is generated indicating the location of the most similar spectra in the data set. The best match of these is extracted from the data set and a traditional mass spectral library search is performed to indicate whether the DotMap analysis was successful.

To investigate the use of DotMap on a complex sample, trimethylsilyl (TMS) derivatized human infant urine containing organic acid metabolites were analyzed. Human infant urine is analyzed in clinical chemistry to diagnose organic acidurias, a class of hereditary inborn errors of metabolism [Chalmers and Lawson 1982]. These metabolic disorders often cause mental retardation and death if not diagnosed and treated, if treatment is available

[Chalmers and Lawson 1982]. Although inborn errors of metabolism individually are rare, collectively small molecule diseases (amino acid diseases, organic acid diseases, primary lactic acidosis, galactosemia, and urea cycle diseases) occur at a rate of 24.4 in 100,000 live births as reported in a recent study in British Columbia, Canada, from 1969 - 1996 [Applegarth et al. 2000]. A common characteristic in the analyses of organic acid disorders is that the aciduria is usually indicated by a large increase in concentration of one or more organic acids in urine [Chalmers and Lawson 1982]. Organic acid analysis in urine is complex even in normal subjects due to the large variations of compounds and concentrations because many organic acids are of exogenous origin [Chalmers and Lawson 1982]. An algorithm such as DotMap can provide a potential benefit in analyses such as these.

DotMap is employed herein to elucidate the location in trimethylsilyl (TMS) derivatized human infant urine of 12 analytes of interest, most of which are commonly found in normal urine. These compounds include the TMS derivatives of lactic acid, pyruvic acid, fumaric acid, adipic acid, suberic acid, 5-oxoproline (pyroglutamic acid), isocitric acid, vanillic acid, homovanillic acid, vanilmandelic acid, 4-hydroxyphenyllactic acid, and m-hippurate. One analyte, vanillic acid (TMS), was located by DotMap and found to be partially overlapped with four interfering compounds. The region containing vanillic acid (TMS) was then subjected to parallel factor analysis (PARAFAC) to deconvolute the pure component signals. Analyte peak deconvolution via PARAFAC has been shown to provide accurate, quantitative results for overlapped analytes in GC x GC-TOFMS data [Sinha et al. 2004a; Sinha et al. 2004b]. PARAFAC is used here to finish the analysis that DotMap begins, adding confidence and additional information to assist in the confirmation of the desired analyte's presence, by providing the pure component mass

spectrum and chromatographic profiles that are required for subsequent quantification.

7.2 Theory

The DotMap algorithm is an adaptation to GC x GC-TOFMS data of mass spectral similarity search algorithms that are based on the dot product of weighted mass spectra [Stein and Scott 1994; Stein 1999]. These previous algorithms were developed and applied to identify sample mass spectra by comparing their mass spectral similarity to collected spectra in a database. The DotMap algorithm calculates the dot product of the scaled, weighted, and normalized analyte mass spectral signal with the mass spectral signal in the data set for each point in the GC x GC plane as follows:

$$\left(\frac{m\sqrt{A_d}}{\sum m\sqrt{A_d}}\right)\bullet\left(\frac{m\sqrt{A_u}}{\sum m\sqrt{A_u}}\right) \tag{7.1}$$

where A_u is the abundance of m/z signals in the "user" spectrum of interest, A_d are the abundances of m/z signals at each point in the two-dimensional GC x GC "data" space, and m is the vector containing m/z values used for weighting the signals. The dot product is denoted by the symbol • and is defined mathematically as:

$$\mathbf{a}\bullet\mathbf{b} = \|\mathbf{a}\|\ \|\mathbf{b}\|\cos\theta \tag{7.2}$$

where **a** and **b** are vectors of identical length, $\|\mathbf{a}\|$ denotes the norm of the vector, and θ represents the angle between the vectors. The dot product gives the projection of **a** (data spectrum) in the direction of **b** (user spectrum). In the case of DotMap, where **a** and **b** are both column vectors of the same length, the dot product is also given by:

$$\mathbf{a}\bullet\mathbf{b} = \mathbf{a}^{T} * \mathbf{b} \tag{7.3}$$

where $\mathbf{a}^T$ denotes the transpose of $\mathbf{a}$ and $*$ indicates vector multiplication. Thus, for DotMap, the more similar the two spectra are, the larger the result of this dot product. Two completely different spectra should theoretically have a dot product of zero, however random noise contributes a small amount to a dot product resulting in a non-zero "baseline" in the raw DotMap results. The algorithm process is shown pictorially in Figure 7.1. The GC x GC-TOFMS data is baseline corrected by subtracting from each point in the data set a scalar quantity, *c*, given by

$$c = \tilde{\mathbf{x}} + 3\sigma_{\mathbf{x}} \qquad (7.4)$$

where $\tilde{\mathbf{x}}$ denotes the median of the vector, $\mathbf{x}$, that contains the minimum value of the GC x GC plane at each *m/z*, and $\sigma_{\mathbf{x}}$ is the standard deviation of $\mathbf{x}$. The mass spectral signal intensities in the data set and spectrum of interest are then scaled to the 0.5 power to enhance the signal from small intensity signals relative to the high intensity signals. Weighting the signals by mass improves performance of the algorithm due to higher *m/z* channels containing more critical information for identification as compared to lower *m/z* channels [Stein and Scott 1994]. Using the DotMap algorithm, a matrix containing the values of the dot product with dimensions of a GC x GC chromatogram is generated. A contour plot is then displayed indicating the location of dot product values above a threshold. The threshold giving the most consistent and accurate results for properly identifying the analyte of interest was determined empirically to be 90% of the maximum dot product value above the median value of all dot products in the raw DotMap data. The algorithm then extracts the spectrum from the GC x GC-TOFMS data that corresponds to the location of the maximum dot product value and performs a traditional mass spectral similarity search. This last step gives immediate feedback on the accuracy of the DotMap analysis.

7.3 Experimental

A derivatized sample of human infant urine containing organic acid metabolites was analyzed to test DotMap performance characteristics on a biological sample. A library of standards built by a collaborator at Children's Hospital & Regional Medical Center in Seattle, Washington and the NIST/EPA/NIH Mass Spectral Library (NIST 02) provided the spectra for the DotMap analysis of the urine sample. Prior to analysis, organic acids in the urine were derivatized with hydroxylamine and N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) to make trimethylsilyl (TMS) derivatives [Tanaka et al. 1980a; Tanaka et al. 1980b]. The analysis of the metabolite sample was performed using a LECO Pegasus 4D GC x GC-TOFMS instrument (LECO Corporation, St. Joseph, MI, USA). Column 1 was a 20-m 250-μm i.d. capillary column with a 0.5-μm film of 5% diphenyl/95% dimethyl polysiloxane (DB-5; J & W Scientific, Alltech). Column 2 was a 2-m 180-μm i.d. capillary column with a 0.2-μm film of trifluoropropylmethyl polysiloxane (Rtx-200; Restek Corp., Bellefonte, PA, USA). These columns were joined using a Vu2 union (Restek Corp.). Modulation was performed using cryogenic modulation. The modulator at the head of column 2 concentrated the column 1 effluent during each column 2 separation. A 0.4 second "hot pulse" began each new column 2 run when the cryogenic gas was shut off and the heated air jets (40 °C above the oven temperature) were switched on. The total column 2 run time was 2 s. Ultra high purity helium (0.8 ml/min) was used as the carrier gas. 2 μL of derivatized sample was injected using a 100:1 split. The column 1 oven ramp began at 60 °C with a hold time of 0.25 min then increased at 8 °C/min to 250 °C and held for 10 min. Column 2 was housed in a separate oven and held at a constant 10 °C

higher than the column 1 oven temperature. No mass spectra were collected during the solvent delay for the first 5 min of each run. The transfer line was maintained at 280 °C and the ion source set point was 200 °C. The detector voltage was -1600 V and the filament bias was -70 V. Mass spectra were collected from *m/z* 41 - 625 at 100 spectra/second. Data were then exported as a comma separated value (.csv) file and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing. The GC x GC-TOFMS data in this exported file has each column 2 GC-MS chromatogram strung end-to-end. Once the data is loaded into Matlab, the column 2 GC-MS chromatograms are segmented-off based on modulation frequency and then stacked to create the GC x GC-TOFMS data cube. The time axes are calculated based on the spectral scan rate and the modulation frequency. The complete DotMap algorithm Matlab executable file (.m file) appears in Appendix A. Mass Spectral similarity searches were performed using the NIST MS Search 2.0 (NIST/EPA/NIH Mass Spectral Library; NIST 02) and the comprehensive MS library built a collaborator at Children's Hospital & Regional Medical Center that included many organic acid TMS derivatives relevant in organic aciduria clinical diagnosis. The deconvolution of the vanillic acid (TMS) was performed using PARAFAC initiated by trilinear decomposition (TLD) [Sinha et al. 2004a; Sinha et al. 2004b]. The TLD algorithm was from the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA) and was selected for the advantageous sequencing of the three dimensions of the matrix prior to analysis. The PARAFAC algorithm was from the N-way Toolbox 2.01 [Andersson and Bro 2000].

7.4 Results and Discussion

The GC x GC-TOFMS analysis of derivatized human infant urine exhibits a highly efficient chromatographic separation that uses a large portion of the available peak capacity (Figure 7.2A). The image in Figure 7.2A depicts the GC x GC chromatogram of *m/z* 73, a selective ion characteristic of trimethylsilyl (TMS) derivatives of organic acids. A selected region of Figure 7.2A is shown to demonstrate the high efficiency of the separation (Figure 7.2B). The locations of TMS derivatives of fumaric acid, adipic acid, and 5-oxoproline (pyroglutamic acid) are indicated. These were located using library spectra and the DotMap algorithm.

An analyte, 5-oxoproline (TMS), is used herein as an example to demonstrate implementation of the DotMap algorithm. 5-oxoproline is an organic acid that is potentially indicative of hawkinsinuria, an inherited disorder of tyrosine metabolism [Borden et al. 1992; Wilcken et al. 1981], or 5-oxoprolinuria (glutathione synthetase deficiency) when found at elevated concentrations in urine [Divry et al. 1991; Metanchez et al. 2001]. The algorithm is automated in that given a data set and an analyte library spectrum, the algorithm will generate a contour plot indicating the location of the most probable match, automatically extract a data spectrum from that location, and perform an automated search for a spectral identity match with the NIST MS Search program within the ~ 30 s analysis time. To more fully explain how the algorithm works, the steps within the algorithm are detailed here. Within the algorithm, the library spectrum (*m/z* 41 - 350) of 5-oxoproline (TMS) was scaled, weighted and normalized according to Eq. (7.1). At each point in the GC x GC plane of the data set, there is a complete mass spectrum (*m/z* 41 - 625). A portion of the full mass spectrum (*m/z* 41 - 350) that corresponds to the mass range of interest for 5-oxoproline (TMS) is also scaled,

weighted and normalized according to Eq. (7.1) after baseline correction (Eq. (7.4)). The dot product is then calculated for the analyte spectrum and the spectrum at each point in the GC x GC plane (as illustrated in Figure 7.1). The dot product intensities are related to how similar the two spectra are at each time element in the GC x GC data plane. The raw data matrix from the DotMap analysis of 5-oxoproline (TMS) projected on to the column 1 dimension is shown in Figure 7.3A. The threshold shown as a dotted line was calculated as described in the theory section as 90% of the difference between the maximum point and the median of the DotMap raw data (Figure 7.3A). An analogous figure showing the raw data in the column 2 dimension was omitted for brevity. Only one signal (i.e., analyte peak) is above the threshold in this raw DotMap data indicating that the similarity between the analyte spectrum and the spectra corresponding to this region in the raw data is high and that there is a high degree of selectivity. A contour plot of the DotMap raw data with a single contour line drawn at the threshold indicated in Figure 7.3A is shown in Figure 7.3B. In these analyses, a contour plot is more useful than the projected raw data because the information from the projections on both columns are combined such that the retention time on both axes can be illuminated from only one plot. A spectrum was extracted from the original baseline corrected data matrix at the coordinates obtained for the maximum point in the DotMap raw data and sent directly to the NIST MS Search program where a similarity search was conducted. A head-to-tail comparison of the extracted data spectrum and the 5-oxoproline (TMS) library spectrum indicates a high level of similarity (Similarity = 880, Reverse = 923) (Figure 7.3C). The time to execute the DotMap algorithm and perform the subsequent MS library database search for 5-oxoproline (TMS) was approximately 30 s on

a PC with a 1.6 GHz processor with 1024 MB of RAM, indicating an efficient analysis time.

Eleven additional DotMap analyses were preformed in a similar manner as that described for 5-oxoproline. Analyses were performed on the TMS derivatives of lactic acid, pyruvic acid, fumaric acid, adipic acid, suberic acid, isocitric acid, vanillic acid, homovanillic acid, vanilmandelic acid, 4-hydroxyphenyllactic acid, and m-hippurate. Depending on the analyte, mass ranges used in the analyses ranged from *m/z* 41 - 350 to *m/z* 41 - 435. As anticipated and desired, for each analysis, only one analyte peak was located that resulted in dot products above the calculated threshold. The locations of the desired compounds were all confirmed by high quality mass spectral search results for the associated extracted spectrum (Similarities ≥ 830). An overlay of all 12 contour plots from the analyses is shown in Figure 7.4. By comparing the image of the original data (Figure 7.2A) and the results of the DotMap analyses (Figure 7.4), one can clearly see that the overabundance of information in the data set has been filtered to efficiently give the desired information. It is also apparent that the region containing TMS derivatives of vanillic, homovanillic, isocitric, vanilmandelic, and 4-hydroxyphenyllactic acids is highly complex, yet DotMap is able to successfully locate the desired analytes in this complex region.

Upon closer inspection of the region surrounding TMS derivatized vanillic acid (*m/z* 73), it is apparent that the acid is overlapped with at least three interfering analytes (Figure 7.5). This region was analyzed with the chemometric algorithm PARAFAC to deconvolute the signals. The data set (*m/z* 45 - 350) surrounding vanillic acid (TMS), was baseline corrected with a linear correction at each mass along column 2 prior to analysis. The region was first modeled by TLD with 5 components specified to get an estimate of

the profiles in each dimension. These results were then used to initiate a PARAFAC analysis, which was conducted with unimodal constraints on the column 1 and column 2 dimensions and nonnegative constraints in all dimensions. The column 1 and column 2 profiles are depicted in Figure 7.6A and B. The solid line indicates the signal for the vanillic acid (TMS) (Figure 7.6A and B), whose identity was postulated because of the previous DotMap analysis. The deconvolution was successful and provided reasonable analyte peak shapes. The vanillic acid (TMS) peak profiles are acceptable for subsequent quantification, if desired for the analysis. A head-to-tail comparison of the deconvoluted spectrum of vanillic acid and the library spectrum of the TMS derivatized vanillic acid (TMS) indicates a high level of similarity (Similarity: 874, Reverse: 885), substantiating the results of the DotMap analysis that located the analyte of interest (Figure 7.6C). The signal heights in the higher *m/z* range are depleted in the deconvoluted spectrum because of the differences in ionization of this TOFMS that typically produces spectra with enriched low *m/z* signals.

7.5 Conclusions

An algorithm to locate analytes of interest in GC x GC-TOFMS data was presented, and found to perform efficiently and accurately. DotMap successfully elucidated the locations of 12 derivatized organic acid metabolites in human infant urine. Therefore, the DotMap algorithm is demonstrated to be a sensitive tool that is useful in the analysis of complex mixtures to reveal the location of analytes of interest. This is particularly useful for analytes in samples exhibiting substantial chemical matrix induced retention time shifts. One analyte in the urine sample, vanillic acid (TMS), was located by DotMap, but also exhibited overlap with other compounds in the sample. The region containing vanillic acid (TMS) and four interfering analytes was successfully

deconvoluted by PARAFAC to confirm its presence and provide pure quantifiable pure component profiles. While DotMap is capable of searching for a spectrum of interest in an entire GC x GC-TOFMS chromatogram in an automated fashion, smaller regions of the data can be analyzed to save analysis time if additional information about probable retention time is known. An investigation into DotMap's limit of detection, dynamic range, optimization, response to analyte peak overlap, response to low mass spectral selectivity (e.g., isomer spectra), and potential to indicate classes of compounds for other complex samples will be the basis of future investigations.

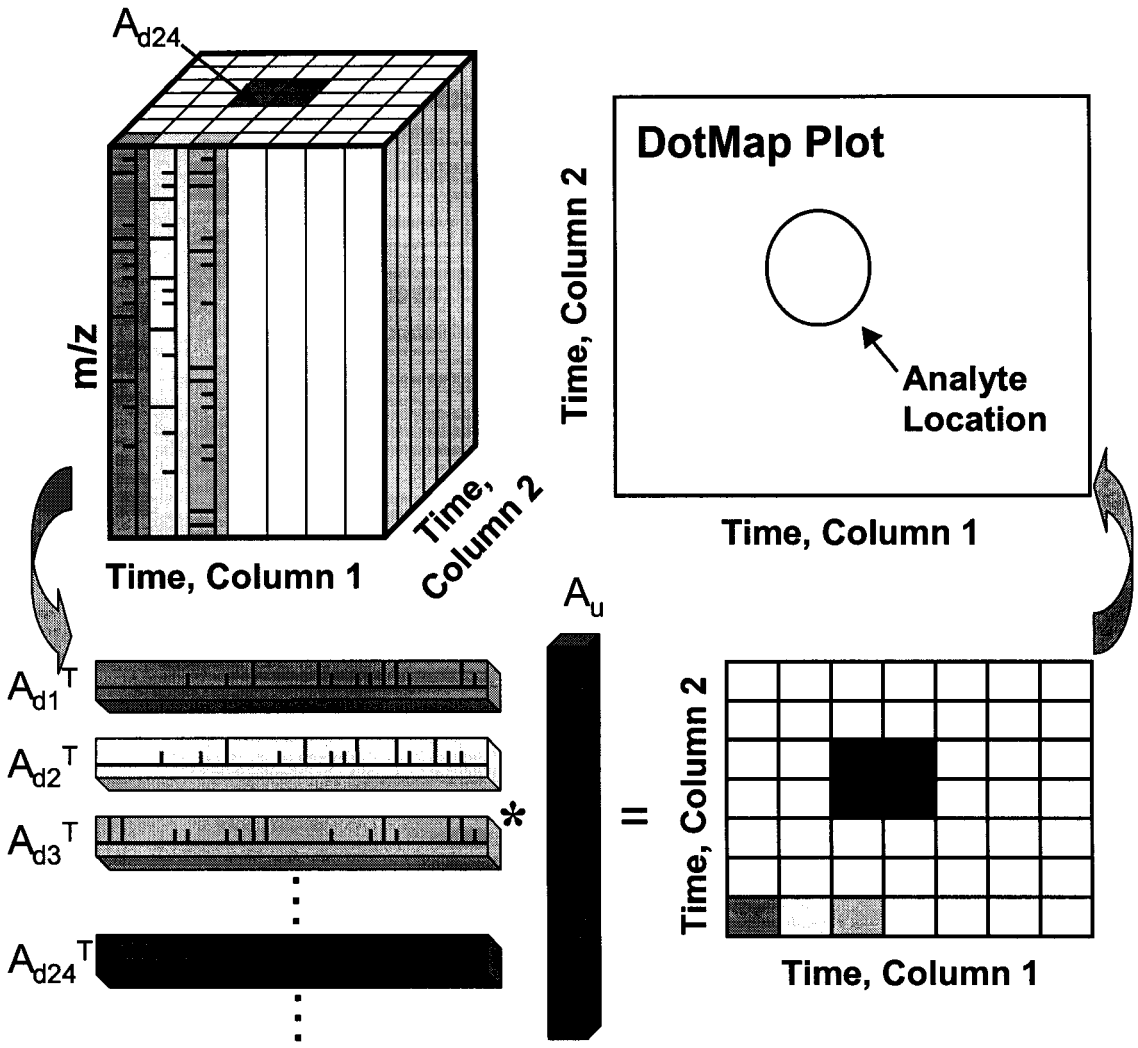


Figure 7.1. Schematic illustrating DotMap algorithm. Dotmap takes a matrix of GC x GC-TOFMS ($\mathbf{A}_d$) "data," and a spectrum of an analyte of interest ($\mathbf{A}_u$) provided by the "user" and generates a 2-D plane and a threshold to designate the location of a compound in the matrix of data that most closely matches the analyte of interest according to Eqs. 7.1 and 7.3. $\mathbf{A}_{d1}$, $\mathbf{A}_{d2}$, and $\mathbf{A}_{d3}$ correspond to the spectra at "data points" 1, 2, and 3, respectively. For example, $\mathbf{A}_{d24}$ corresponds to the 24th spectrum in the data set as shown.

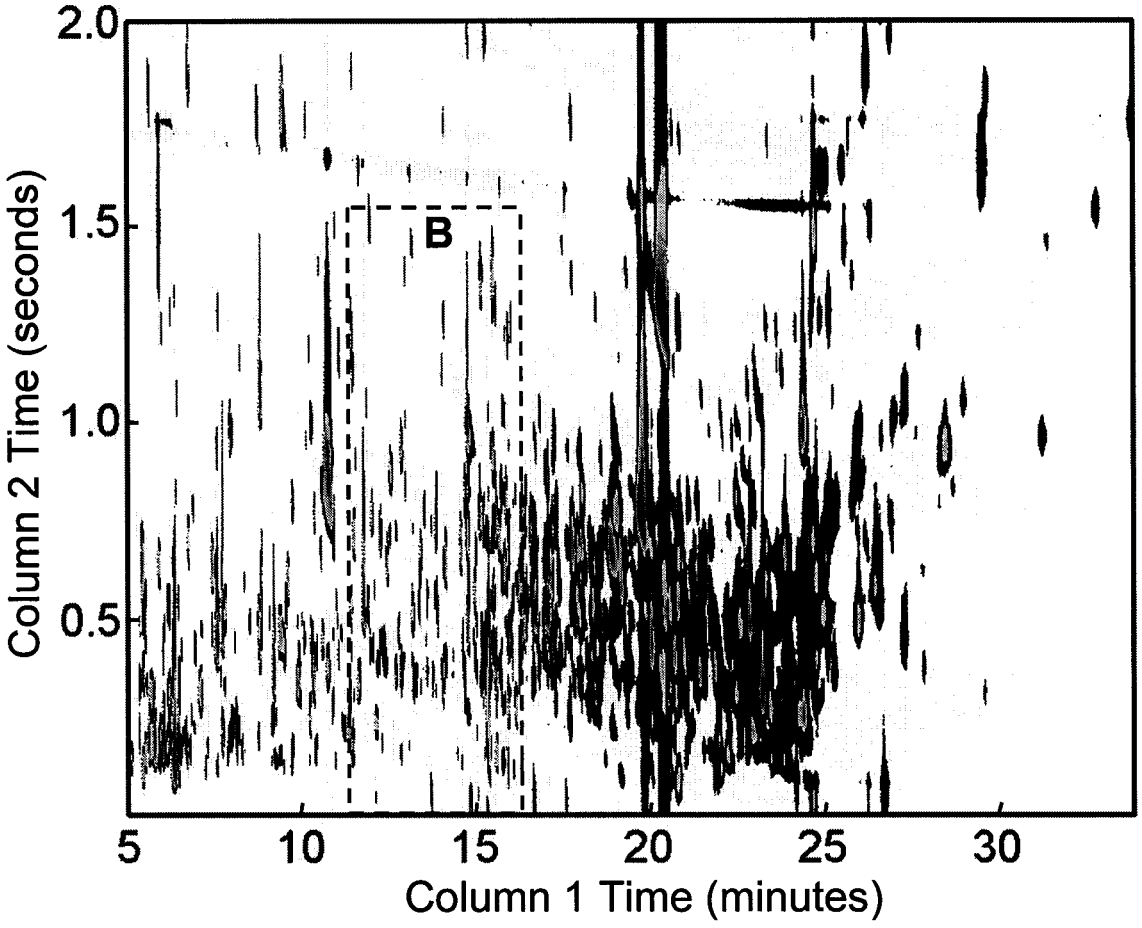


Figure 7.2A. GC x GC-TOFMS analysis of TMS derivatized organic acids in human infant urine. Chromatogram at *m/z* 73 is shown. The region in the box is shown in Figure 7.2B.

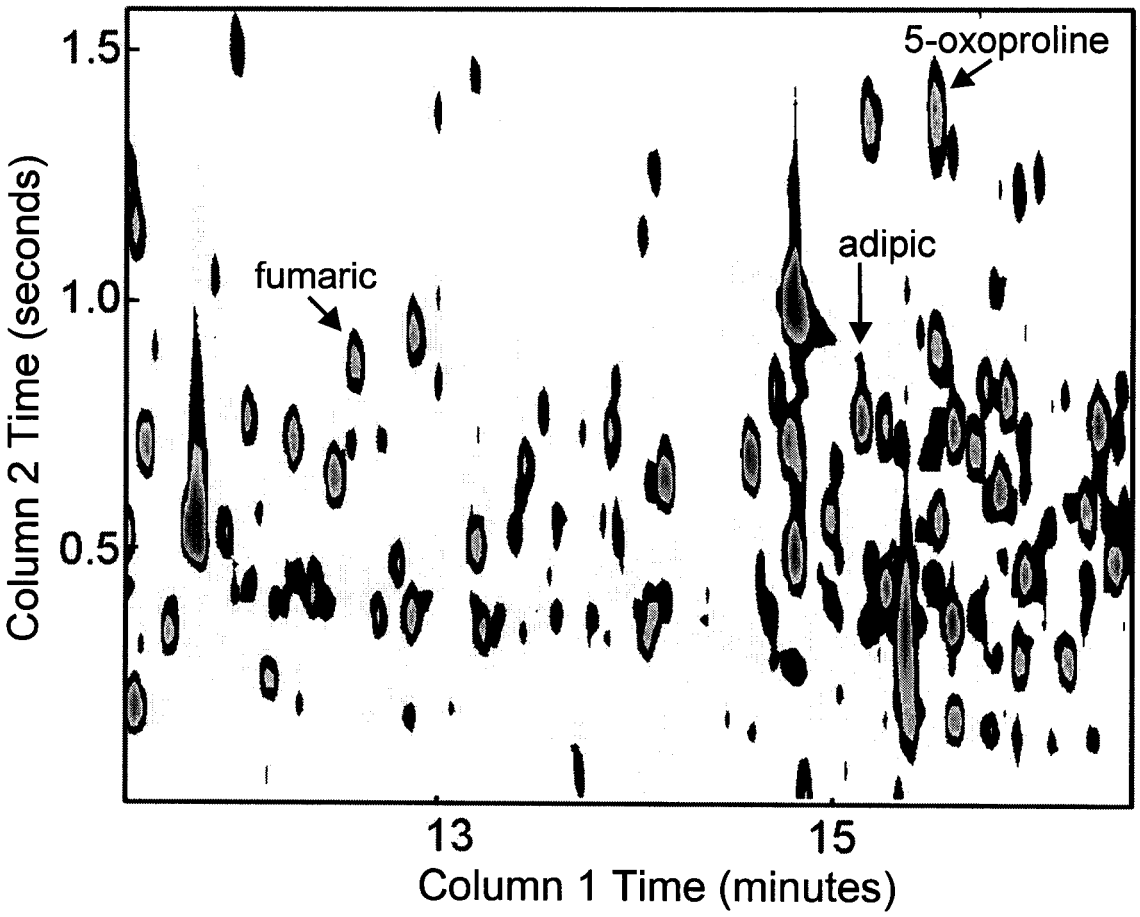


Figure 7.2B. Sub-region of TMS derivatized organic acid analysis. Chromatogram at *m/z* 73 is shown. This region of the chromatogram in Figure 7.2A demonstrates high separation efficiency and sample complexity. The locations of TMS derivatized 5-oxoproline (pyroglutamic acid), fumaric acid and adipic acid are indicated. 5-oxoproline (TMS) is used for DotMap illustration as shown in Figure 7.3.

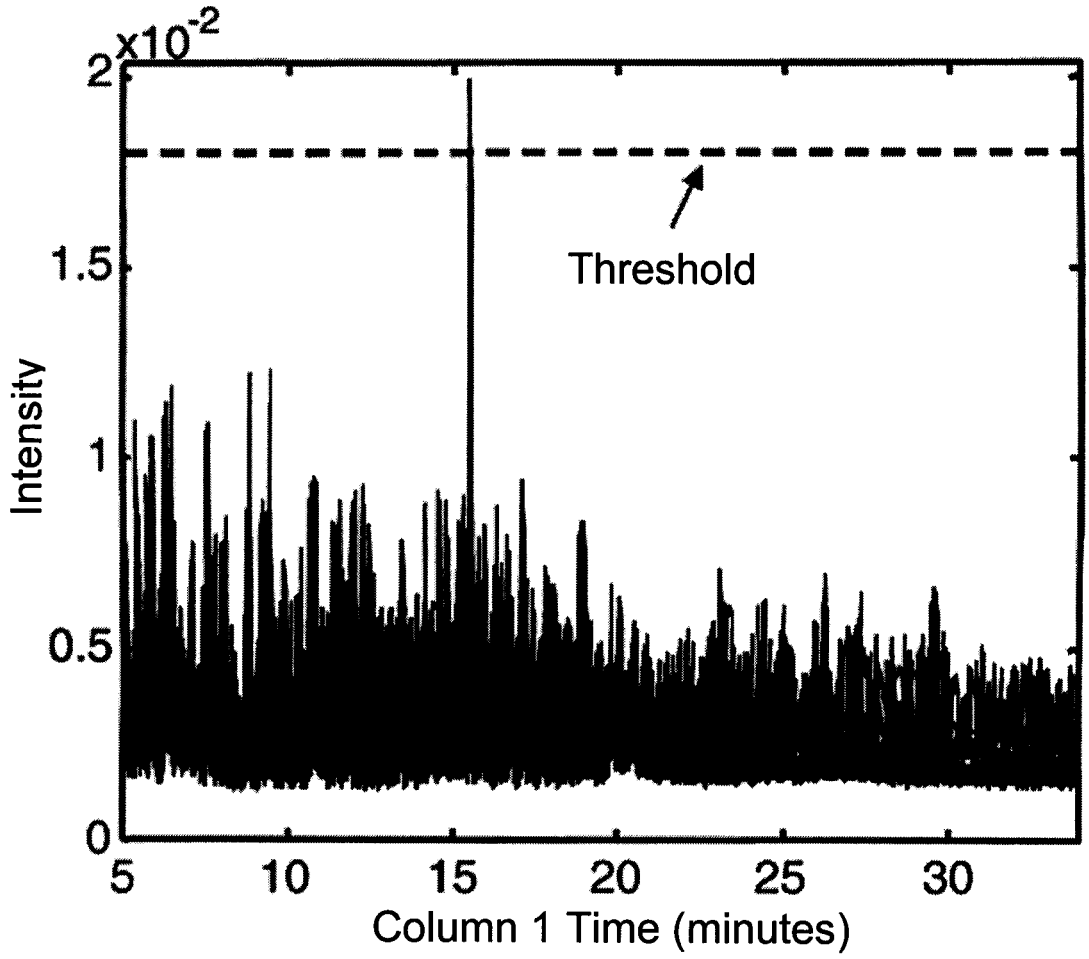


Figure 7.3A. Raw data for Column 1 from the DotMap analysis of 5-oxoproline (TMS). The maximum peak indicated the location of 5-oxoproline on Column 1. The threshold indicated is 90% of the maximum dot product value above the median value of all dot products in the raw DotMap data. The location on Column 2 is found in an analogous fashion, but not shown for brevity.

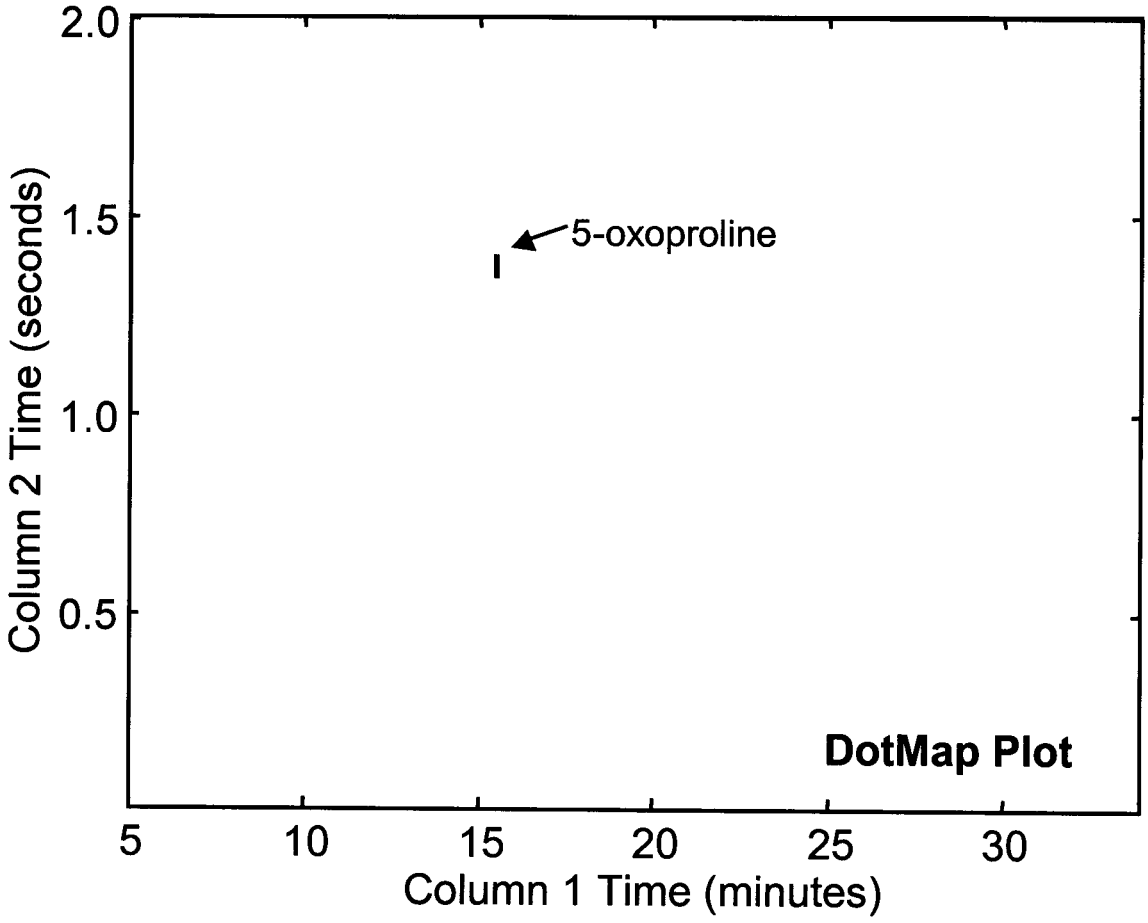


Figure 7.3B. DotMap result for 5-oxoproline (TMS). Location of analyte indicated.

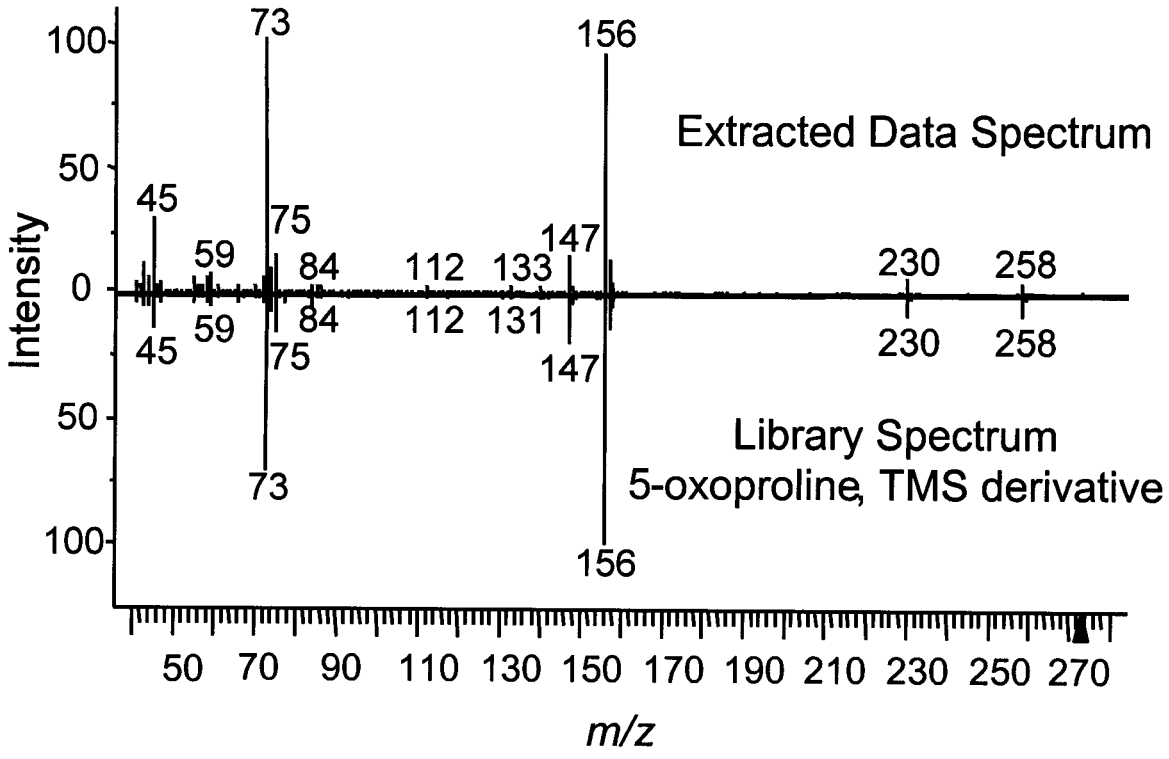


Figure 7.3C. Extracted mass spectrum from the DotMap analysis of 5-oxoproline (TMS). The extracted spectrum was extracted from the maximum point in the DotMap analysis shown in Figure 7.3B. The spectrum was successfully matched to TMS derivatized 5-oxoproline from the NIST library.

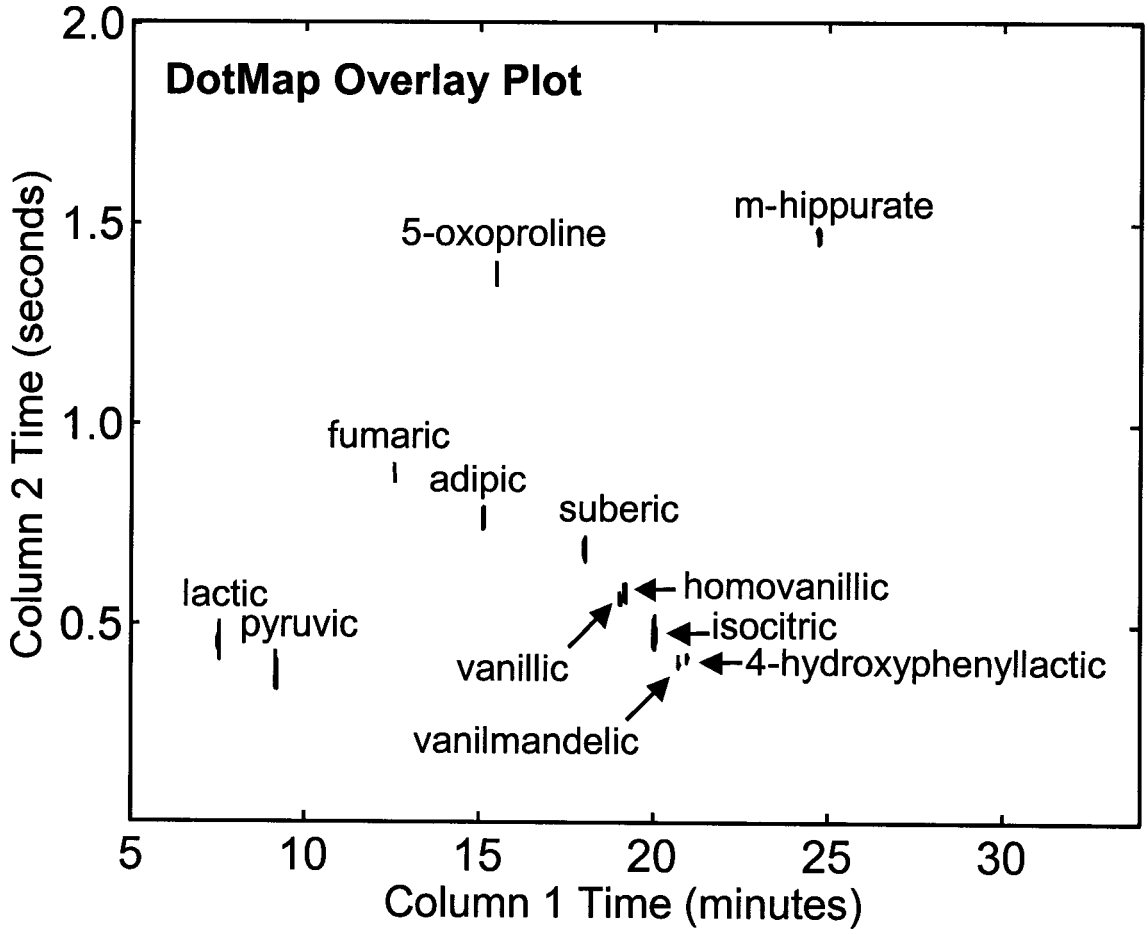


Figure 7.4. Overlays of 12 DotMap analyses. The locations of 12 TMS derivatized organic acids in human infant urine are indicated on the plot. Compare analyte locations to complexity of elution region in Figure 7.2A. Confirmation of analysis was achieved by matching DotMap-extracted spectra to library spectra.

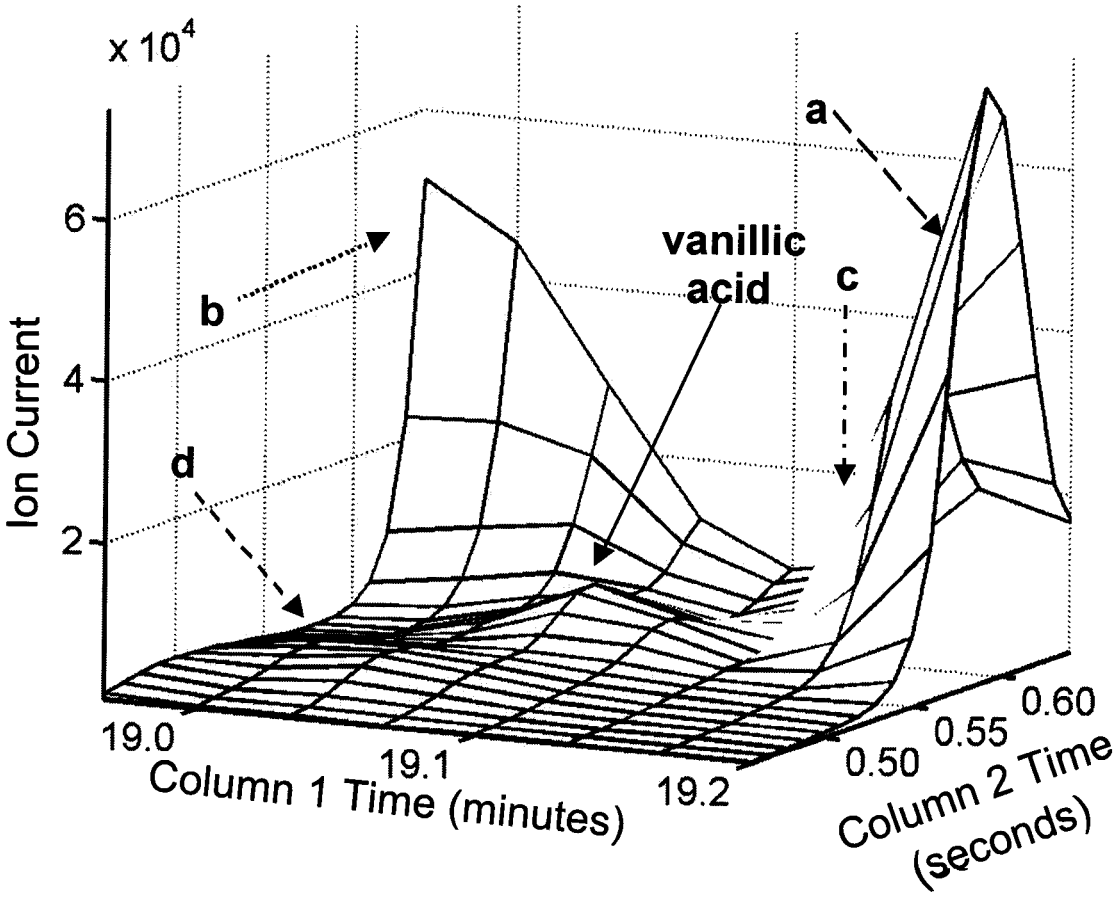


Figure 7.5. 3-D mesh plot of *m/z* 73 of the region surrounding vanillic acid (TMS). Four interfering analytes **a** - **d** are labeled. The line types of the labeling arrows correspond to the line types for the PARAFAC results in Figure 7.6.

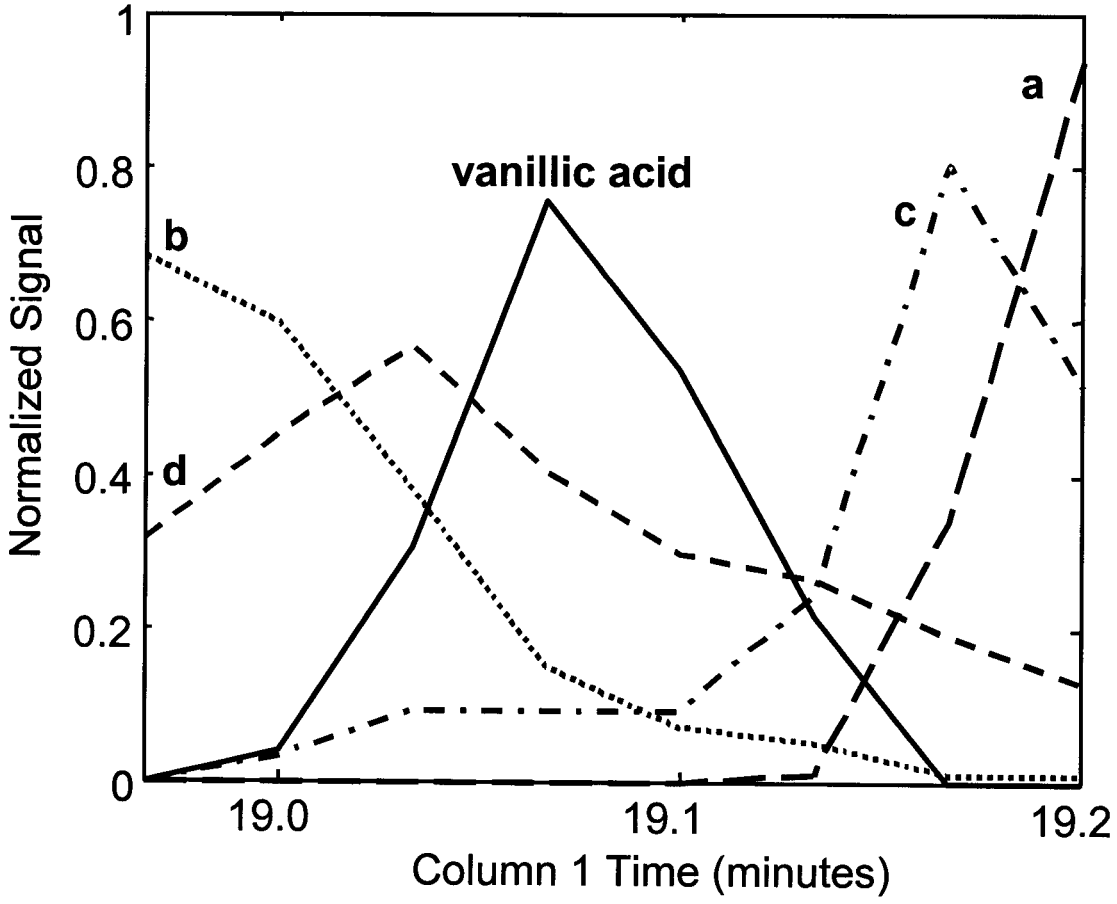


Figure 7.6A. PARAFAC results for column 1 of vanillic acid (TMS) region. Normalized signals of vanillic acid (TMS) and interfering analytes (**a** - **d**) are shown.

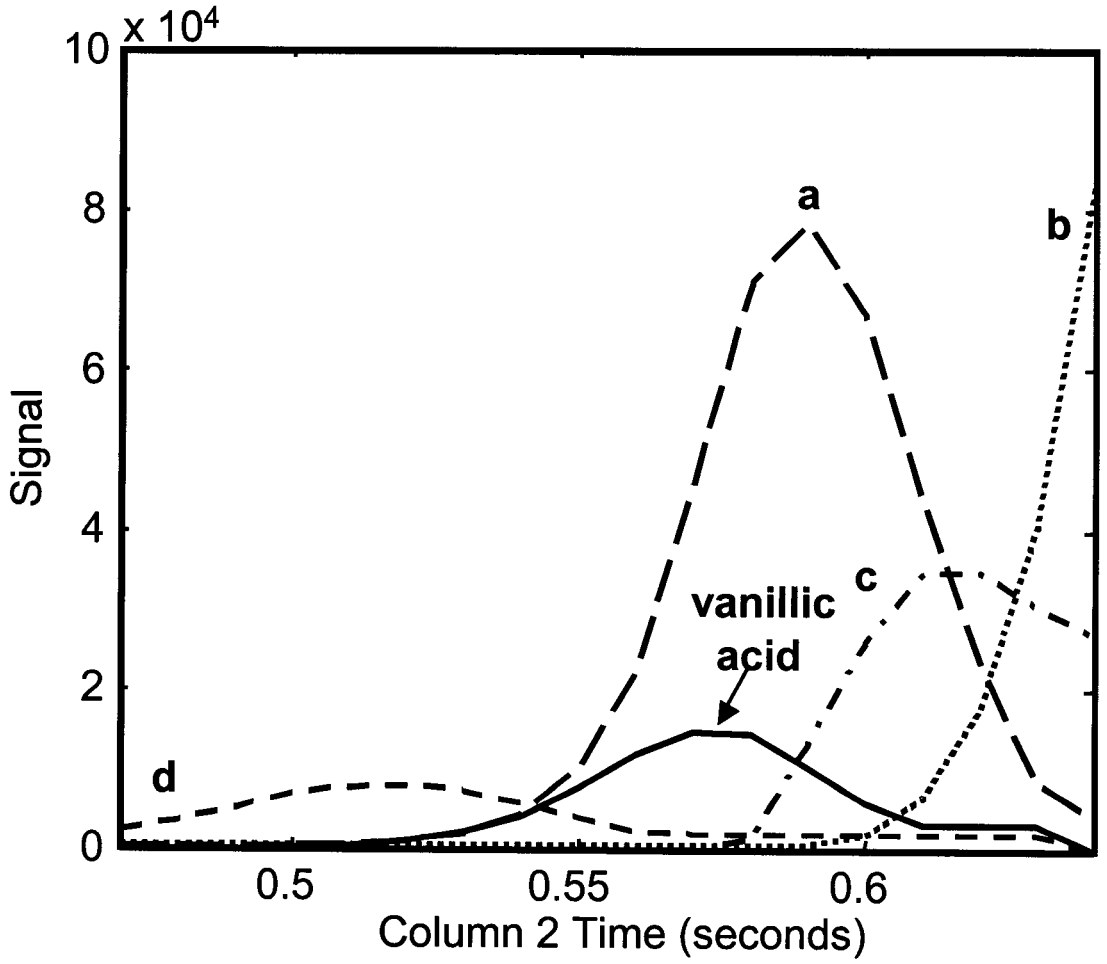


Figure 7.6B. PARAFAC results for column 2 of vanillic acid (TMS) region. Signals including concentration information of vanillic acid (TMS) and interfering analytes (**a** – **d**) are shown.

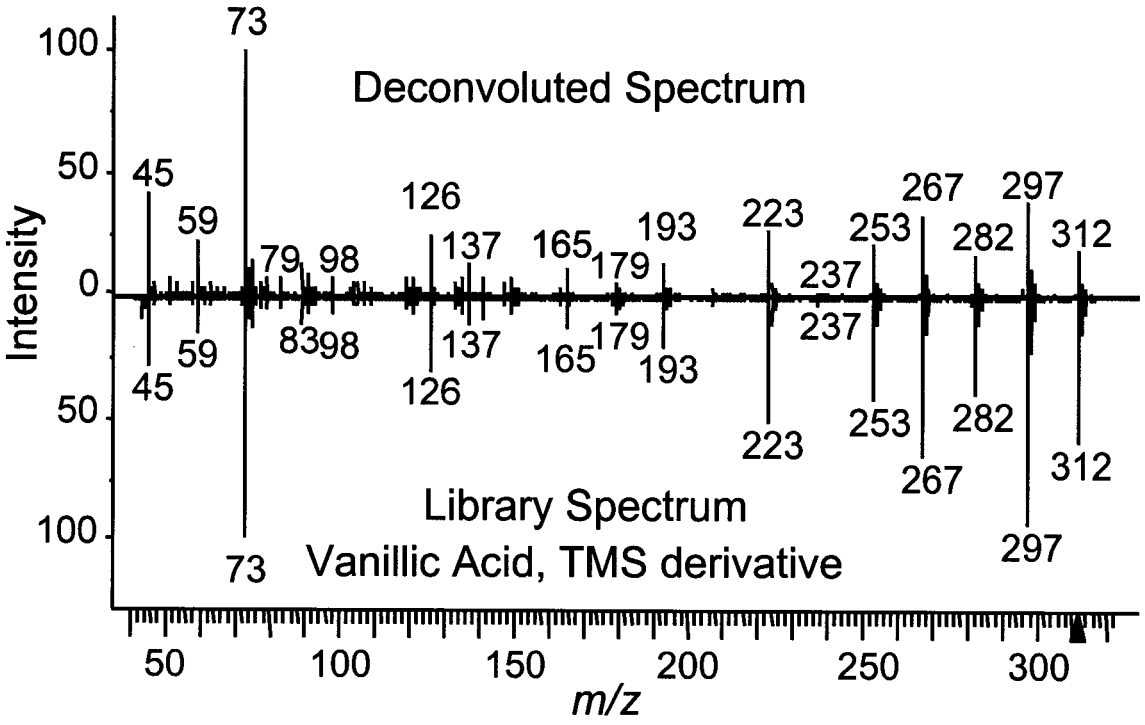


Figure 7.6C. Deconvoluted mass spectrum from PARAFAC model of vanillic acid (TMS) region matched to derivatized vanillic acid (TMS) in the NIST library.

Chapter 8. Final Conclusions and Future Directions

Aspects of GC x GC and GC x GC-TOFMS development have been presented for both valve-based instruments and thermal modulation-based instruments. Most significant, however, is the development of chemometric methodology for signal deconvolution. PARAFAC has been shown in this work to successfully deconvolute overlapping peaks with low mass spectral selectivity and low chromatographic resolution. While a simple comparison of other applicable peak deconvolution techniques was presented, a more thorough investigation is in order to provide a more complete understanding of the benefits and drawbacks to each method. One potentially beneficial development would be to extend AMDIS to work with third-order data, i.e., GC x GC-TOFMS data, in addition to GC-MS data.

Another tool that was presented to simplify the overabundance of GC x GC-TOFMS data was DotMap. This new tool was able to pinpoint the locations of analytes of interest in complex GC x GC-TOFMS data sets. A more thorough investigation of the algorithm would be beneficial to investigate the effects of mass spectra similarity and peak overlap on the performance of the algorithm. In addition, an understanding of the tool's practical limitations, e.g., LOD, is necessary. The possibility to tune the results of DotMap to generate maps of single analytes or class information should also be investigated.

There is still much research to be done to bridge the gap between 2D separations and chemometrics. One major area that has the potential to meet some of these challenges and advance this field is the development of novel algorithms for retention time alignment of 2D data. Although most of the chemometric work to date has been done with GC x GC data, the inherent characteristics of all 2D separations, e.g., LC x CE and CE x CE, make them

amenable to chemometric analysis. Some of the most exciting new and further developments may come by applying chemometric pattern recognition with 2D separations and mass spectrometric detection. Ultimately, the amount of information resulting from these separations is so great that only with the implementation of chemometric techniques will the desired information be garnered efficiently and completely.

BIBLIOGRAPHY

Adahchour, M.; van Stee, L. L. P.; Beens, J.; Vreuls, R. J. J.; Batenburg, M. A. and Brinkman, U. A. Th. (2003) "Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection for the Trace Analysis of Flavour Compounds in Food," *Journal of Chromatography A*, 1019: 157-172.

Adams, M. A.; Chen, Z.; Landman, P. and Colmer, T. D. (1999) "Simultaneous Determination by Capillary Gas Chromatography of Organic Acids, Sugars, and Sugar Alcohols in Plant Tissue Extracts as Their Trimethylsilyl Derivatives," *Analytical Biochemistry*, 266: 77-84.

Andersson, C. A. and Bro, R. (2000) "The N-Way Toolbox for Matlab," *Chemometrics and Intelligent Laboratory Systems*, 52: 1-4.

Applegarth, D. A.; Toone, J. R. and Lowry, R. B. (2000) "Incidence of Inborn Errors of Metabolism in British Columbia, 1969-1996," *Pediatrics*, 105: Art. No. e10.

Beens, J.; Adahchour, M.; Vreuls, R. J. J.; van Altena, K. and Brinkman, U. A. Th. (2001a) "Simple, Non-Moving Modulation Interface for Comprehensive Two-Dimensional Gas Chromatography," *Journal of Chromatography A*, 919: 127-132.

Beens, J.; Dalluge, J.; Adahchour, M.; Vreuls, R. J. J. and Brinkman, U. A. Th. (2001b) "Moving Cryogenic Modulator for Hte Comprehensive Two-Dimensional Gas Chromatography (GC x GC) of Surface Water Contaminants," *Journal of Microcolumn Separations*, 13: 134-140.

Bertsch, W. (2000) "Two-Dimensional Gas Chromatography. Concepts, Instrumentation, and Applications - Part 2: Comprehensive Two-Dimensional Gas Chromatography," *Journal of High Resolution Chromatography*, 23: 167-181.

Booksh, K. S. and Kowalski, B. R. (1994) "Theory of Analytical Chemistry," *Analytical Chemistry*, 66: 782A-791A.

Booksh, K. S.; Lin, Z. H.; Wang, Z. Y. and Kowalski, B. R. (1994) "Extension of Trilinear Decomposition Method with an Application to the Flow Probe Sensor," *Analytical Chemistry*, 66: 2561-2569.

Borden, M.; Hom, J.; Leslie, J.; Sweetman, L.; Nyhan, W. L.; Fleisher, L.; Nadler, H.; Lewis, D. and Scott, C. R. (1992) "Hawkinsinuria in Two Families," *American Journal of Medical Genetics*, 44: 52-56.

Bro, R. (1997) "PARAFAC. Tutorial and Applications," *Chemometrics and Intelligent Laboratory Systems*, 38: 149-171.

Bruckner, C. A.; Prazen, B. J. and Synovec, R. E. (1998) "Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis," *Analytical Chemistry*, 70: 2796-2804.

Bushey, M. M. and Jorgenson, J. W. (1990a) "Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography of Proteins," *Analytical Chemistry*, 62: 161-167.

Bushey, M. M. and Jorgenson, J. W. (1990b) "Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography/Capillary Zone Electrophoresis," *Analytical Chemistry*, 62: 978-984.

Chalmers, R. A. and Lawson, A. M. (1982) Organic Acids in Man: Analytical Chemistry, Biochemistry and Diagnosis of the Organic Acidurias. London: Chapman and Hall.

Cotter, R. J. (ed). (1994) Time-of-Flight Mass Spectrometry. (ACS Symposium Series, vol 549), Washington, D.C.: American Chemical Society.

Cotter, R. J. (1997) Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research. Washington, D.C.: American Chemical Society.

Dallüge, J.; Beens, J. and Brinkman, U. A. Th. (2003) "Comprehensive Two-Dimensional Gas Chromatography: A Powerful and Versatile Analytical Tool," *Journal of Chromatography A*, 1000: 69-108.

Dallüge, J.; van Rijn, M.; Beens, J.; Vreuls, R. J. J. and Brinkman, U. A. Th. (2002a) "Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection Applied to the Determination of Pesticides in Food Extracts," *Journal of Chromatography A*, 965: 207-217.

Dallüge, J.; van Stee, L. L. P.; Xu, X.; Williams, J.; Beens, J.; Vreuls, R. J. J. and Brinkman, U. A. Th. (2002b) "Unravelling the Composition of Very Complex Samples by Comprehensive Gas Chromatography Coupled to Time-of -Flight Mass Spectrometry: Cigarette Smoke," *Journal of Chromatography A*, 974: 169-184.

Dallüge, J.; Vreuls, R. J. J.; Beens, J. and Brinkman, U. A. Th. (2002c) "Optimization and Characterization of Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection (GC x GC-TOFMS)," *Journal of Separation Science*, 25: 201-214.

de Geus, H. J.; Baycan-Keller, R.; Oehme, M.; de Boer, J. and Brinkman, U. A. Th. (1998) "Determination of Enantiomer Ratios of Bornane Congeners in Biological Samples Using Heart-Cut Multidimensional Gas Chromatography," *Journal of High Resolution Chromatography*, 21: 39-46.

de Geus, H. J.; Schelvis, A.; de Boer, J. and Brinkman, U. A. Th. (2000) "Comprehensive Two-Dimensional Gas Chromatography with a Rotating Thermal Desorption Modulator and Independently Temperature-Programmable Columns," *Journal of High Resolution Chromatography*, 23: 189-196.

de Geus, H. J.; Aidos, I.; de Boer, J.; Luten, J. B. and Brinkman, U. A. Th. (2001) "Characterisation of Fatty Acids in Biological Oil Samples Using Comprehensive Multidimensional Gas Chromatography," *Journal of Chromatography A*, 910: 95-103.

de Koning, S.; Lach, G.; Linkerhagner, M.; Loscher, R.; Tablack, P. H. and Brinkman, U. A. Th. (2003) "Trace-Level Determination of Pesticides in Food Using Difficult Matrix Introduction-Gas Chromatography-Time-of-Flight Mass Spectrometry," *Journal of Chromatography A*, 1008: 247-252.

Debonneville, C. and Chaintreau, A. (2004) "Quantitation of Suspected Allergens in Fragrances Part II. Evaluation of Comprehensive Gas Chromatography-Conventional Mass Spectrometry," *Journal of Chromatography A*, 1027: 109-115.

Dietz, E. A. (1996) "Shifting of Gas Chromatographic Retention Times Due to Solvent Effects - a Study Using Sulfur Chemiluminescence Detection," *HRC-Journal of High Resolution Chromatography*, 19: 485-491.

Dimandja, J.-M. D.; Stanfill, S. B.; Grainger, J. and Donald G. Patterson, J. (2000) "Application of Comprehensive Two-Dimensional Gas Chromatography (GC x GC) to the Qualitative Analysis of Essential Oils," *Journal of High Resolution Chromatography*, 23: 208-214.

Dimandja, J.-M. D. (2003) "A New Tool for the Optimized Analysis of Complex Volatile Mixtures: Comprehensive Two-Dimensional Gas Chromatography / Time-of-Flight Mass Spectrometry," *American Laboratory*, 35: 42-53.

Divry, P.; Roulaud-Parrot, F.; Dorche, C.; Zabot, M. T.; Contraire, B.; Hagenfeldt, L. and Larsson, A. (1991) "5-Oxoprolinuria (Glutathione Synthetase Deficiency): A Case with Neonatal Presentation and Rapid Fatal Outcome," *Journal of Inherited Metabolic Disease*, 14: 341-344.

Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C. and Duffield, A. M. (1976) "Extraction of Mass-Spectra Free of Background and Neighboring Component Contributions from Gas Chromatography Mass Spectrometry Data," *Analytical Chemistry*, 48: 1368-1375.

Faber, K.; Lober, A. and Kowalski, B. R. (1997) "Analytical Figures of Merit for Tensorial Calibration," *Journal of Chemometrics*, 11: 419-461.

Faber, N. M. (2002) "Towards a Rehabilitation of the Generalized Rank Annihilation Method (GRAM)," *Analytical and Bioanalytical Chemistry*, 372: 683-687.

Faber, N. K. M.; Bro, R. and Hopke, P. K. (2003) "Resent Developments in CANDECOMP/PARAFAC Algorithms: A Critical Review," *Chemometrics and Intelligent Laboratory Systems*, 65: 119-137.

Fiehn, O.; Kopka, J.; Trethewey, R. N. and Willmitzer, L. (2000) "Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry," *Analytical Chemistry*, 72: 3573-3580.

Focant, J.-F.; Sjödin, A. and Patterson, D. G., Jr. (2003) "Qualitative Evaluation of Thermal Desorption-Programmable Temperature Vaporization-Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry for the Analysis of Selected Halogenated Contaminants," *Journal of Chromatography A*, 1019: 143-156.

Foley, J. P. and Dorsey, J. G. (1983) "Equations for Calcualtion of Chromatrographic Figures of Merit for Ideal and Skewed Peaks," *Analytical Chemistry*, 55: 730-737.

Fraga, C. G.; Prazen, B. J. and Synovec, R. E. (2000a) "Comprehensive Two-Dimensional Gas Chromatography and Chemometrics for the High-Speed Quantitative Analysis of Aromatic Isomers in a Jet Fuel Using the Standard Addition Method and an Objective Retention Time Alignment Algorithm," *Analytical Chemistry*, 72: 4154-4162.

Fraga, C. G.; Prazen, B. J. and Synovec, R. E. (2000b) "Enhancing the Limit of Detection for Comprehensive Two-Dimensional Gas Chromatography (GC x GC) Data Using Bilinear Chemometric Analysis," *Journal of High Resolution Chromatography*, 3: 215-224.

Fraga, C. G.; Bruckner, C. A. and Synovec, R. E. (2001a) "Increasing the Number of Analyzable Peaks in Comprehensive Two-Dimensional Separations through Chemometrics," *Analytical Chemistry*, 73: 675-683.

Fraga, C. G.; Prazen, B. J. and Synovec, R. E. (2001b) "Objective Data Alignment and Chemometric Analysis of Comprehensive Two-Dimensional Separations with Peak Shifting on Both Dimensions," *Analytical Chemistry*, 73: 5833-5840.

Frysinger, G. S.; Gaines, R. B. and Ledford, E. B., Jr. (1999) "Quantitative Determination of BTEX and Total Aromatic Compounds in Gasoline by Comprehensive Two-Dimensional Gas Chromatography (GC x GC)," *Journal of High Resolution Chromatography*, 22: 195-200.

Frysinger, G. S. and Gaines, R. B. (2001) "Separation and Identification of Petroleum Biomarkers by Comprehensive Two-Dimensional Gas Chromatography," *Journal of Separation Science*, 24: 87-96.

Giddings, J. C. (1984) "Two-Dimensional Separations: Concept and Promise," *Analytical Chemistry*, 56: 1258A-1270A.

Giddings, J. C. (1987) "Concepts and Comparisons in Multidimensinal Separation," *Journal of High Resolution Chromatography & Chromatography Communications*, 10: 319-323.

Giddings, J. C. (1990) "Use of Multiple Dimensions in Analytical Separations," In Cortes, H. (ed.) Multidimensional Chromatography: Techniques and Applications. New York: Marcel Dekker, Inc., 1-27.

Giddings, J. C. (1991) Unified Separation Science. New York: Wiley.

Grob, R. L. (1985) "Theory of Gas Chromatography," In Grob, R. L. (ed.) Modern Practice of Gas Chromatography, second ed. New York: John Wiley & Sons, 49-115.

Gui, M.; Rutan, S. C. and Agbodjan, A. (1995) "Kinetic Detection of Overlapped Amino Acids in Thin-Layer Chromatography with a Direct Trilinear Decomposition Method," *Analytical Chemistry*, 67: 3293-3299.

Guiochon, G.; Gonnord, M. F.; Siouffi, A. and Zakaria, M. (1982) "Study of the Performances of Thin-Layer Chromatography. 7. Spot Capacity in Two-Dimensional Thin-Layer Chromatography," *Journal of Chromatography*, 250: 1-20.

Halket, J. M.; Przyborowski, A.; Stein, S. E.; Mallard, W. G.; Down, S. and Chalmers, R. A. (1999) "Deconvolution Gas Chromatography/Mass Spectrometry of Urinary Organic Acids - Potential for Pattern Recognition and Automated Identification of Metabolic Disorders," *Rapid Communications in Mass Spectrometry*, 13: 279-284.

Harju, M. and Haglund, P. (2001) "Comprehensive Two-Dimensional Gas Chromatography (GC x GC) of Atropisomeric PCBs, Combining a Narrow Bore Beta-Cyclodextrin Column and a Liquid Crystal Column," *Journal of Microcolumn Separations*, 13: 300-305.

Harju, M.; Bergman, A.; Olsson, M.; Roos, A. and Haglund, P. (2003a) "Determination of Atropisomeric and Planar Polychlorinated Biphenyls, Their Enantiomeric Fractions and Tissue Distribution in Grey Seals Using Comprehensive 2D Gas Chromatography," *Journal of Chromatography A*, 1019: 127-142.

Harju, M.; Danielsson, C. and Haglund, P. (2003b) "Comprehensive Two-Dimensional Gas Chromatography of the 209 Polychlorinated Biphenyls," *Journal of Chromatography A*, 1019: 111-126.

Harris, R. A.; Yang, A.; Stein, R. C.; Lucy, K.; Brusten, L.; Herath, A.; Parekh, R.; Waterfield, M. D.; O'Hare, M. J.; Neville, M. A.; Page, M. J. and Zvelebil, M. J. (2002) "Cluster Analysis of an Extensive Human Breast Cancer Cell Line Protein Expression Map Database," *Proteomics*, 2: 212-223.

Holland, L. A. and Jorgenson, J. W. (2000) "Characterization of a Comprehensive Two-Dimensional Anion Exchange-Perfusive Reversed Phase Liquid Chromatography System for Improved Separations of Peptides," *Journal of Microcolumn Separations*, 12: 371-377.

Hua, R. X.; Li, Y. Y.; Liu, W.; Zheng, J. C.; Wei, H. B.; Wang, J. H.; Lu, X.; Kong, H. W. and Xu, G. W. (2003) "Determination of Sulfur-Containing Compounds in Diesel Oils by Comprehensive Two-Dimensional Gas Chromatography with a Sulfur Chemiluminescence Detector," *Journal of Chromatography A*, 1019: 101-109.

Johnson, K.; De Juan, A. and Rutan, S. C. (1999) "Three-Way Data Analysis of Pollutant Degradation Profiles Monitored Using Liquid Chromatography-Diode Array Detection," *Journal of Chemometrics*, 13: 331-341.

Johnson, K. J.; Prazen, B. J.; Olund, R. K. and Synovec, R. E. (2002) "GC x GC Temperature Programming Requirements to Produce Bilinear Data for Chemometric Analysis," *Journal of Separation Science*, 25: 297-303.

Johnson, K. J. and Synovec, R. E. (2002) "Pattern Recognition of Jet Fuels: Comprehensive GC x GC with ANOVA-Based Feature Selection and Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, 60: 225-237.

Johnson, K. J.; Wright, B. W.; Jarman, K. H. and Synovec, R. E. (2003) "A High-Speed Peak Matching Algorithm for Retention Time Alignment of Gas Chromatographic Data for Chemometric Analysis," *Journal of Chromatography A*, 996: 141-155.

Kaiser, M. A. (1985) "High-Resolution Gas Chromatography," In Grob, R. L. (ed.) Modern Practice of Gas Chromatography. New York: John Wiley & Sons, 159-185.

Kinghorn, R. M. and Marriott, P. J. (2000) "Design and Implementation of Comprehensive Gas Chromatography with Cryogenic Modulation," *Journal of High Resolution Chromatography*, 23: 245-252.

Korytar, P.; Leonards, P. E. G.; de Boer, J. and Brinkman, U. A. Th. (2002) "High-Resolution Separation of Polychlorinated Biphenyls by Comprehensive Two-Dimensional Gas Chromatography," *Journal of Chromatography A*, 958: 203-218.

Korytar, P.; van Stee, L. L. P.; Leonards, P. E. G.; de Boer, J. and Brinkman, U. A. Th. (2003) "Attempt to Unravel the Composition of Toxaphene by Comprehensive Two-Dimensional Gas Chromatography with Selective Detection," *Journal of Chromatography A*, 994: 179-189.

Ledford, E. B., Jr.; Phillips, J. B.; Xu, J.; Gaines, R. B. and Blomberg, J. (1996) "Ordered Chromatograms: A Powerful Methodology in Gas Chromatography," *American Laboratory*, 28: 22-25.

Ledford, E. B., Jr. and Billesbach, C. (2000) "Jet-Cooled Thermal Modulator for Comprehensive Multidimensional Gas Chromatography," *Journal of High Resolution Chromatography*, 23: 202-204.

Lee, A. L.; Bartle, K. D. and Lewis, A. C. (2001) "A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography," *Analytical Chemistry*, 73: 1330-1335.

Lewis, A. C.; Carslaw, N.; Marriott, P. J.; Kinghorn, R. M.; Morrison, P.; Lee, A. L.; Bartle, K. D. and Pilling, M. J. (2000) "A Larger Pool of Ozone-Forming Carbon Compounds in Urban Atmospheres," *Nature*, 45: 778-781.

Li, S. and Gemperline, P. J. (1993) "Eliminating Complex Eigenvectors and Eigenvalues in Multiway Analyses Using the Direct Triinear Decomposition Method," *Journal of Chemometrics*, 7: 77-88.

Liu, Z. and Phillips, J. B. (1991) "Comprehensive Two-Dimensional Gas Chromatography Using an On-Column Thermal Modulator Interface," *Journal of Chromatographic Science*, 29: 227-231.

Lin, Z.; Booksh, K. S.; Burgess, L. and Kowalski, B. R. (1994) "Second-Order Fiber Optic Heavy Metal Sensor Employing Second-Order Tensorial Calibration," *Analytical Chemistry*, 66: 2552.

Liu, Z.; Patterson, D. G., Jr. and Lee, M. L. (1995) "Geometric Approach to Factor Analysis for the Estimation of Orthogonality and Practical Peak Capacity in Comprehensive Two-Dimensional Separations," *Analytical Chemistry*, 67: 3840-3845.

Liu, Z. and Lee, M. L. (2000) "Comprehensive Two-Dimensional Separations Using Microcolumns," *Journal of Microcolumn Separations*, 12: 241-254.

Lorber, A.; Faber, K. and Kowalski, B. R. (1997) "Net Analyte Signal Calculation in Multivariate Calibration," *Analytical Chemistry*, 69: 1620-1626.

Lu, X.; Cai, J.; Kong, H.; Wu, M.; Hua, R.; Zhao, M.; Liu, J. and Xu, G. (2003) "Analysis of Cigarette Smoke Condensates by Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry: I. Acidic Fraction," *Analytical Chemistry*, 75: 4441-4451.

Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V. and Zagluin, V. A. (1973) "The Mass-Reflectron, a New Nonmagnetic Time-of-Flight Mass Spectrometer with High Resolution," *Soviet Physics-JETP*, 37: 45-48.

Marengo, E.; Leardi, R.; Robotti, E.; Righetti, P. G.; Antonucci, F. and Cecconi, D. (2003) "Application of Three-Way Principal Component Analysis to the Evaluation of Two-Dimensional Maps in Proteomics," *Journal of Proteome Research*, 2: 351-360.

Marriott, P. J. and Kinghorn, R. M. (1997) "Longitudinally Modulated Cryogenic System. A Generally Applicable Approach to Solute Trapping and Mobilization in Gas Chromatography," *Analytical Chemistry*, 69: 2582-2588.

Marriott, P. J.; Shellie, R.; Fergeus, J.; Ong, R. and Morrison, P. (2000) "High Resolution Essential Oil Analysis by Using Comprensive Gas Chromatographic Methodology," *Flavour and Fragrance Journal*, 15: 225-239.

Marriott, P. J. and Shellie, R. (2002) "Principles and Applications of Comprehensive Two-Dimensional Gas Chromatography," *Trends in Analytical Chemistry*, 21: 573-583.

Metanchez, D.; Rabier, D.; Mokhtari, M.; Durrmeyer, X.; Megnet, A.; Larsson, A. and Saudubray, J.-M. (2001) "5-Oxoprolinuria: A Cause of Neonatal Metabolic Acidosis," *Acta Paediatrica*, 90: 827-828.

Michels, D. A.; Hu, S.; Schoenherr, R. M.; Eggertson, M. J. and Dovichi, N. J. (2002) "Fully Automated Two-Dimensional Capillary Electrophoresis for High Sensitivity Protein Analysis," *Molecular & Cellular Proteomics*, 1: 69-74.

Moore, A. W., Jr. and Jorgenson, J. W. (1995) "Rapid Comprehensive Two-Dimensional Separations of Peptides Via RPLC-Optically Gated Capillary Zone Electrophoresis," *Analytical Chemistry*, 67: 3448-3455.

Murphy, R. E.; Schure, M. R. and Foley, J. P. (1998) "Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography," *Analytical Chemistry*, 70: 1585-1594.

Phillips, J. B. and Xu, J. (1995) "Comprehensive Multi-Dimensional Gas Chromatography," *Journal of Chromatography A*, 703: 327-334.

Phillips, J. B. and Beens, J. (1999) "Comprhensive Two-Dimensional Gas Chromatography: A Hyphenated Method with Strong Coupling between the Two Dimensions," *Journal of Chromatography A*, 856: 331-347.

Phillips, J. B.; Gaines, R. B.; Blomberg, J.; van der Wielen, F. W. M.; Dimandja, J. M.; Green, V.; Granger, J.; Patterson, D.; Racovalis, L.; de Geus, H. J.; de Boer, J.; Haglund, P.; Lipsky, J.; Sinha, V. and Ledford, E. B., Jr. (1999) "A Robust Thermal Modulator for Comprehensive Two-Dimensional Gas Chromatography," *Journal of High Resolution Chromatography*, 22: 3-10.

Poe, R. B. and Rutan, S. C. (1993) "Effects of Resolution, Peak Ratio and Sampling Frequency in Diode-Array Fluorescence Detection in Liquid Chromatography," *Analytica Chimica Acta*, 283: 845-853.

Prazen, B. J.; Bruckner, C. A.; Synovec, R. E. and Kowalski, B. R. (1999a) "Enhanced Chemical Analysis Using Parallel Column Gas Chromatography with Single-Detector Time-of-Flight Mass Spectrometry and Chemometric Analysis," *Analytical Chemistry*, 71: 1093-1099.

Prazen, B. J.; Bruckner, C. A.; Synovec, R. E. and Kowalski, B. R. (1999b) "Second Order Chemometric Standardization for High Speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GC x GC Data," *Journal of Microcolumn Separations*, 11: 97-107.

Prazen, B. J.; Johnson, K. J.; Weber, A. and Synovec, R. E. (2001) "Two-Dimensional Gas Chromatography and Trilinear Partial Least Squares for the Quantitative Analysis of Aromatic and Naphthene Content in Naphtha," *Analytical Chemistry*, 73: 5677-5682.

Quigley, W. W. C.; Fraga, C. G. and Synovec, R. E. (2000) "Comprehensive LC x GC for Enhanced Headspace Analysis," *Journal of Microcolumn Separations*, 12: 160-166.

Sánchez, E. and Kowalski, B. R. (1986) "Generalized Rank Annihilation Factor Analysis," *Analytical Chemistry*, 58: 496-499.

Sánchez, E. and Kowalski, B. R. (1988) "Tensorial Calibration: II. Second-Order Calibration," *Journal of Chemometrics*, 2: 265-280.

Sánchez, E. and Kowalski, B. R. (1990) "Tensorial Resolution: A Direct Trilinear Decomposition," *Journal of Chemometrics*, 4: 29-45.

Seeley, J. V.; Kramp, F. and Hicks, C. J. (2000) "Comprehensive Two-Dimensional Gas Chromatography Via Differential Flow Modulation," *Analytical Chemistry*, 72: 4346-4352.

Seeley, J. V.; Kramp, F. J.; Sharpe, K. S. and Seeley, S. K. (2002) "Characterization of Gaseous Mixtures of Organic Compounds with Dual-Secondary Column Comprehensive Two-Dimensional Gas Chromatography (GC x 2GC)," *Journal of Separation Science*, 25: 53-59.

Shellie, R.; Marriott, P. J. and Cornwell, C. (2001a) "Application of Comprehensive Two-Dimensional Gas Chromatography (GC x GC) to the Enantioselective Analysis of Essential Oils," *Journal of Separation Science*, 24: 823-830.

Shellie, R.; Marriott, P. J. and Morrison, P. (2001b) "Concepts and Preliminary Observations on the Triple-Dimensional Analysis of Complex Volatile Samples by Using GC x GC-TOFMS," *Analytical Chemistry*, 73: 1336-1344.

Shellie, R.; Mondello, L.; Marriott, P. and Dugo, G. (2002) " Characterisation of Lavender Essential Oils by Using Gas Chromatography–Mass Spectrometry with Correlation of Linear Retention Indices and Comparison with Comprehensive Two-Dimensional Gas Chromatography," *Journal of Chromatography A*, 970: 225-234.

Shellie, R. A. and Marriott, P. J. (2003) "Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Analysis of Pelargonium Graveolens Essential Oil Using Rapid Scanning Quadrupole Mass Spectrometry," *Analyst*, 128: 879-883.

Shellie, R. A.; Marriott, P. J. and Huie, C. W. (2003) "Comprehensive Two-Dimensional Gas Chromatography (GC x GC) and GC x GC-Quadrupole MS Analysis of Asian and American Ginseng," *Journal of Separation Science*, 26: 1185-1192.

Shoenmakers, P.; Marriott, P. J. and Beens, J. (2003) "Nomenclature and Conventions in Comprehensive Multidimensional Chromatography," *LC-GC Europe*, 16: 335-338.

Sinha, A. E.; Johnson, K. J.; Prazen, B. J.; Lucas, S. V.; Fraga, C. G. and Synovec, R. E. (2003a) "Comprehensive Two-Dimensional Gas Chromatography (GC x GC) of Volatile and Semi-Volatile Components Using a Diaphragm Valve-Based Instrument," *Journal of Chromatography A*, 983: 195-204.

Sinha, A. E.; Prazen, B. J.; Fraga, C. G. and Synovec, R. E. (2003b) "Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection: Instrumentation and Figures of Merit," *Journal of Chromatography A*, 1019: 79-87.

Sinha, A. E.; Fraga, C. G.; Prazen, B. J. and Synovec, R. E. (2004a) "Trilinear Chemometric Analysis of Two-Dimensional Comprehensive Gas Chromatography-Time-of-Flight Mass Spectrometry Data," *Journal of Chromatography A*, 1027: 269-277.

Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J. and Synovec, R. E. (2004b) "Multivariate Selectivity as a Metric for Evaluating GC x GC/TOFMS Data Subjected to Chemometric Peak Deconvolution," *Journal of Chromatography A*, Accepted for publication.

Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M. and Synovec, R. E. (2004c) "An Algorithm for Locating Analytes of Interest Based on Mass Spectral Similarity in GC x GC-TOFMS Data," *Journal of Chromatography A*, Submitted for publication.

Sinha, A. E.; Prazen, B. J. and Synovec, R. E. (2004d) "Trends in Chemometric Analysis of Comprehensive Two-Dimensional Separations," *Analytical and Bioanalytical Chemistry*, 378: 1948-1951.

Slonecker, P. J., Li, X.; Ridgway, T. H. and Dorsey, J. G. (1996) "Information Orthogonality of Two-Dimensional Chromatographic Separations," *Analytical Chemistry*, 68: 682.

Stein, S. E. and Scott, D. R. (1994) "Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification," *Journal of the American Society for Mass Spectrometry*, 5: 859-866.

Stein, S. E. (1999) "An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry," *Journal of the American Society for Mass Spectrometry*, 10: 770-781.

Stephens, W. E. (1946) "A Pulsed Mass Spectrometer with Time Dispersion," *Physical Review*, 69: 691.

Synovec, R. E.; Fraga, C. G.; Prazen, B. J.; Sinha, A. E. and Johnson, K. J. (2003a) "GC x GC/TOFMS with Chemometric Data Analysis," In: The First International Symposium on Comprehensive Multidimensional Gas Chromatography. Volendam, the Netherlands.

Synovec, R. E.; Prazen, B. J.; Johnson, K. J.; Fraga, C. G. and Bruckner, C. A. (2003b) "Chemometric Analysis of Comprehensive Two-Dimensional Separations," In Brown, P. R., Grushka, E. (eds.) Advances in Chromatography. New York: Marcel Dekker, Inc., 1-42.

Tanaka, K.; West-Dull, A.; Hine, D. G.; Lynn, T. B. and Lowe, T. (1980a) "Gas-Chromatographic Method of Analysis for Urinary Organic Acids. I. Retention Indices of 155 Metabolically Important Compounds," *Clinical Chemistry*, 26: 1839-1846.

Tanaka, K.; West-Dull, A.; Hine, D. G.; Lynn, T. B. and Lowe, T. (1980b) "Gas-Chromatographic Method of Analysis for Urinary Organic Acids. II. Description of The Procedure, and Its Application to Diagnosis of Patients with Organic Acidurias," *Clinical Chemistry*, 26: 1847-1853.

Truong, T. T.; Marriott, P. J.; Porter, N. A. and Leeming, R. (2003) "Application of Comprehensive Two-Dimensional Gas Chromatography to the Quantification of Overlapping Faecal Sterols," *Journal of Chromatography A*, 1019: 197-210.

van Deursen, M.; Beens, J.; Reijenga, J.; Lipman, P.; Cramers, C. and Blomberg, J. (2000) "Group-Type Identification of Oil Samples Using Comprehensive Two-Dimensional Gas Chromatography Couples to a Time-of-Flight Mass Spectrometer (GC x GC-Tof)," *Journal of High Resolution Chromatography*, 23: 507-510.

van Mispelaar, V. G.; Tas, A. C.; Smilde, A. K.; Schoenmakers, P. J. and van Asten, A. C. (2003) "Quantitative Analysis of Target Components by Comprehensive Two-Dimensional Gas Chromatography," *Journal of Chromatography A*, 1019: 15-29.

Venkatramani, C. J.; Xu, J. and Phillips, J. B. (1996) "Separation Orthogonality in Temperature-Programmed Comprehensive Two-Dimensional Gas Chromatography," *Analytical Chemistry*, 68: 1486-1492.

Veriotti, T. and Sacks, R. (2003) "Characterization and Quantitative Analysis with GC/TOFMS Comparing Enhanced Separation with Tandem-Column Stop-Flow GC and Spectral Deconvolution of Overlapping Peaks," *Analytical Chemistry*, 75: 4211-4216.

Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M. and Yates, J. R. (2002) "Analysis of Quantitative Proteomic Data Generated Via Multidimensional Protein Identification Technology," *Analytical Chemistry*, 74: 1650-1657.

Washburn, M. P.; Ulaszek, R. R. and Yates, J. R. (2003) "Reproducibility of Quantitative Proteomic Analyses of Complex Biological Mixtures by Multidimensional Protein Identification Technology," *Analytical Chemistry*, 75: 5054-5061.

Welthagen, W.; Schnelle-Kries, J. and Zimmermann, R. (2003) "Search Criteria and Rules for Comprehensive Two-Dimensional Gas Chromatography - Time-of-Flight Mass Spectrometry Analysis of Airborne Particulate Matter," *Journal of Chromatography A*, 1019: 223-249.

Wilcken, B.; Hammond, J. W.; Howard, N.; Bohane, T.; Hocart, C. and Halpern, B. (1981) "Hawkinsinuria: A Dominantly Inherited Defect of Tyrosine Metabolism with Severe Effects in Infancy," *The New England Journal of Medicine*, 305: 865-869.

Wilson, B. E.; Sánchez, E. and Kowalski, B. R. (1989) "An Improved Algorithm for the Generalized Rank Annihilation Method," *Journal of Chemometrics*, 3: 493-498.

Wold, S. (1995) "Chemometrics: What Do We Mean with It, and What Do We Want from It?," *Chemometrics and Intelligent Laboratory Systems*, 30: 109-115.

Zrostlikova, J.; Hajslova, J. and Cajka, T. (2003) "Evaluation of Two-Dimensional Gas Chromatography-Time- of-Flight Mass Spectrometry for the Determination of Multiple Pesticide Residues in Fruit," *Journal of Chromatography A*, 1019: 173-186.

APPENDIX A: Matlab Algorithms for DotMap Analysis

```
function [match] = dotmap(three_d, vect, start, masses2)
%     [match] = dotmap(three_d, vect, start)
%
%     A. Sinha 12/03

[m,n,o] = size(three_d);
%generate vector of m/z values for weighting
for i=0:o-1
   masses(:,i+1)=start+i;
   i+1;
end

masses=masses';
%weight and normalize analyte spectrum
vect = vect.^0.5;
vect=vect.*masses;
vect = vect/sum(vect);

%weight and normalize dataset
basecor=median(min(min(three_d)));
if basecor >= 300
   base=basecor+3*std(min(min(three_d)));
   three_d=three_d-base;
   three_d=abs(three_d);
   clear base
end
dotter=three_d;
clear basecor
dotter = dotter.^0.5;
dotter = reshape(dotter,(m*n),o);
i=0;
for i=0:(m*n-1)
   dotter(i+1,:) = dotter(i+1,:).*masses';
   i+1;
end
%clear masses
pack
%normalize dataset
```

```
i=0;
dotter=reshape(dotter,m,n,o);
tic = squeeze(sum(dotter,3));

for i=0:o-1
   dotter(:,:,i+1) = real(dotter(:,:,i+1)./tic);
   i+1;
end
clear tic
pack
%reshape and take dot product
dotter = reshape(dotter,(m*n),o);
if nargin==3
   dotter = dotter*vect;
else
   dotter = dotter(:,masses2-start+1)*vect(masses2-start+1,:);
end
match = real(reshape(dotter,m,n,1));
clear vect dotter
pack
%define the threshold for drawing a contour line and plot the results
thresh=0.90*(max(max(match))-
median(median(match)))+(median(median(match)));
figure;
contour(match,[thresh thresh],'k')

[I,J] = IND2SUB(size(match),find(match==max(max(match))));

nistfile02(['MS_' num2str(I) '_' num2str(J)], masses , squeeze(three_d(I,J,:)));
```

```
function [ ]=nistfile02(name,masses,data);
%function to generate NIST format mass spec file for database matching
%[ ]=nistfile02(name,masses,data);
%Opens NIST02 for search results

lng=length(masses);
str=['Name: ' name '\nNumpeaks: ' num2str(lng)];

 for i=1:lng
    str=[str '\n' num2str(masses(i)) ' ' num2str(data(i))];
 end

 fid=fopen(['C:\Amanda\Searches\' name], 'w');
fprintf(fid,str);
fclose(fid);
clear fid str;
fid=fopen(['C:\Amanda\Searches\locator.txt'], 'w');
str=['C:\Amanda\Searches\' name];
res=findstr(str, '\');
res=[1 res length(str)];
cstr=[ ];
for i=1:length(res)-2
   cstr=[cstr str(res(i):res(i+1)-1) '\'];
end
cstr=[cstr str(res(i+1):res(i+2))];
cstr=[cstr ' Overwrite\n23 62789'];
fprintf(fid,cstr);
fclose(fid);

dos('C:\NIST02\MSSEARCH\nistms$.exe /insturment/par=4');
```

Vita

Amanda Elizabeth Moses Sinha was born on August 28, 1977 in Roanoke, Virginia. She moved with her family at the age of six to Phoenix, Arizona where she acclimated to the hot weather and sunshine. After graduating from Xavier College Preparatory, she attended Harvey Mudd College in Claremont, California where she received her B.S. in Chemistry in 1999. She then moved to Seattle, Washington to begin graduate school in the Department of Chemistry at the University of Washington. After dabbling in atmospheric chemistry, she decided to study analytical chemistry for her degree, but not before she met her husband, Parikhit Ricky Sinha, who does more than just dabble in atmospheric chemistry. They were married on February 3, 2002 and she received her MS degree that June. As she has now earned her PhD in 2004, and as she never quite acclimated to all the rain and clouds, it is time to move on to bigger and better things.